RADIO FUNDAMENTALS

RADIO FUNDAMENTALS

by ARTHUR LEMUEL ALBERT, M.S., E.E.

Professor of Communication Engineering, Oregon State College

NEW YORK TORONTO LONDON

McGRAW-HILL BOOK COMPANY, INC.

1948

PREFACE

As indicated by the title "Radio Fundamentals," this book covers the basic acoustical and electrical principles of radio. The book is designed primarily for beginning students, for radio technicians, and for radio amateurs rather than for advanced radio engineers.

Progress in radio has been rapid, particularly in recent years, and a sincere effort has been made to reflect this progress in "Radio Fundamentals." Because progress is so dependent on a thorough knowledge of fundamentals, these fundamentals have been stressed. Numerous applications have been included, however, as illustrations, and to make the treatment interesting.

Although the discussions in this book are nonmathematical and at a level for beginning students, the explanations are sound fundamentally, and in accordance with modern viewpoints and standards. In preparing the manuscript, a distinction has been carefully maintained between simplicity of presentation and statements that are vague and technically incorrect.

It is assumed that the readers have some training in elementary electrical theory. As a review, much of Chaps. II, III, and IV is devoted to the basic electrical theories of greatest importance in a study of radio. These chapters are quite condensed, and in some instances supplemental reading in books such as Timbie's "Basic Electricity for Communications" or the author's "Electrical Fundamentals of Communication" may be advisable. There may be a tendency to omit these chapters; before this is done, however, they should be examined carefully because they contain much important material on radio and present information that is needed in understanding the rest of the book.

A number of the illustrations used, and several examples and problems, have been taken from the author's "Electrical Fundamentals of Communication." This was done for convenience and also because these illustrations and the material have proved to be very satisfactory for the specific purposes used.

In preparing the manuscript and in securing illustrations, many

organizations were of direct assistance. In particular, the author is indebted to the following: The American Radio Relay League; The American Telephone and Telegraph Company; Bell Laboratories Record; Bell System Technical Journal; Bell Telephone Laboratories; Blaw-Knox Company; Bliley Electric Company; Brush Development Co.; Centralab Division of Globe Union, Inc.; Collins Radio Co.; The Daven Company; Federal Telephone and Radio Corp.; General Electric Co.; General Radio Co.; International Telephone and Telegraph Corp.; E F. Johnson Co.; Meissner Division of Maguire Industries, Inc.; National Broadcasting Company; Radio Corporation of America; Sperry Gyroscope Company; D. Van Nostrand Company; Western Electric Co.; and Westinghouse Electric Corporation. Many other organizations also were of assistance.

Once again it is a pleasure for the author to acknowledge the assistance of his wife, who typed the manuscript most skillfully and made innumerable helpful suggestions.

ARTHUR LEMUEL ALBERT

CORVALLIS, ORE.
September, 1948

CONTENTS

CHAPTER I

FUNDAMENTALS OF ACOUSTICS

Radio is used for the transmission of information and entertainment. A radio channel may be operated by a telegraph key, by a teletypewriter, or by facsimile equipment, or it may be used for television. Of the many applications of radio that can be enumerated, probably it is used more extensively for transmitting speech and music than for other purposes.

Sometimes the speech or music is reproduced electrically from records, but often a radio system is used to transmit speech or music that originates as sound waves in a studio. Thus a radio-transmitting system often is actuated directly by sound waves, and the distant radio-receiving set reproduces the sound waves at the location of the listener.

Those interested in mastering the subject of radio completely must have a knowledge of sound. The studio or auditorium in which sounds are created affects the intelligibility and quality of the signals that operate the radio microphone. Also, the intelligibility and quality of the sound waves are affected by the room in which they are reproduced by the loudspeaker in the radio-receiving set.

Because of its basic importance, sound, and the related phenomena of speech, music, and the acoustics of studios and auditoriums, will be considered in this chapter.

Sound.—As the term is usually employed, **sound** is a wave motion of air particles. Thus when the key of a piano is struck, a taut wire is caused to vibrate, air particles are set in motion, and sound is radiated. Or, when a loudspeaker diaphragm vibrates, air particles are set in motion, and sound is radiated.

Sounds have several characteristics as follows:

Frequency of Sound.—To refer again to a piano, the various wires that are used to produce the musical tones are of different types, lengths, and sizes, and they are stretched so that they vibrate with different basic frequencies. These vibrations produce sound waves that correspond to the vibrations of the piano

strings. When the wire is moving in a given direction, the air particles ahead of it are compressed slightly, and the air immediately behind it is rarefied slightly. This produces a succession of sound waves in the air, as shown in Fig 1

A large radio studio with splayed ceilings and polycylindrical walls for diffusing and scattering the program sounds The curved surfaces are of plywood or similar material (*Courtesy of Tele-Tech and radio station WNEW*)

The motion of the air particles and of a sound wave is represented usually by the so-called "sine wave" of Fig 2. This wave shows how the positions of the air particles vary with time. The designations + and − are purely arbitrary. As time passes (or progresses), the air particles move back and forth about their normal position of rest

The **frequency** of a sound wave is the number of complete vibrations occurring per second. A **cycle** is one complete set of positive and negative values, such as the one shown in Fig. 2. In

the discussion of speech and music, the characteristic that has been defined here as frequency often is called **pitch.** Thus it is usual to speak of the pitch of a musical tone. A tone of high frequency, or high pitch, has many thousand complete cycles, or vibrations, per second.

FIG. 1.—When a vibrating wire or other object moves back and forth in air, sound waves are produced. Regions where the lines are close together represent compression of the air particles and an increase in air pressure. Regions where the lines are few represent expansion of the air and a decrease in pressure. This shows a simple form of vibration. The strings of a musical instrument usually vibrate in a more complex manner.

Intensity of Sound.—If the vibration of a wire or other sound-producing object is relatively feeble, then the motion of the air particles also will be feeble, and the sound produced will be weak. The **intensity** of a sound wave is determined by the amplitude of vibration of the air particles. This amplitude is small; even for very intense sounds the air particles move back and forth distances of only a few thousandths of an inch.

Energy of Sound Waves.—When the key of a piano is struck by the finger, energy is imparted to the piano wire, and some of this energy is transferred to adjacent air particles. As the sound wave is transmitted through the air, energy is passed from one air particle to another. Thus a sound wave contains energy, although comparatively speaking the amount of energy in a sound wave is very low (page 16). Nevertheless, even with this small amount of energy, the familiar principle of energy conservation applies; the energy in a soundwave cannot be "destroyed," but can be converted from sound energy into energy of some other type, such as heat energy.

FIG. 2.—As a sound wave passes through the air, the air particles move back and forth. The plus and minus signs indicate that the motions are in opposite directions about the position of rest.

Energy in the form of sound waves is transmitted through air at about 1125 feet per second under ordinary conditions of air temperature and pressure. Sound

waves cannot be transmitted through a vacuum (greatly reduced air pressure) because there are insufficient air particles to produce the wave motion.

Reflection and Absorption of Sound.—The application of the principle of conservation of energy to sound is of assistance in explaining the phenomena associated with the reflection of sound.

To study sound reflection, suppose that a loudspeaker is sending out sound waves as indicated in Fig. 3. When the sound waves

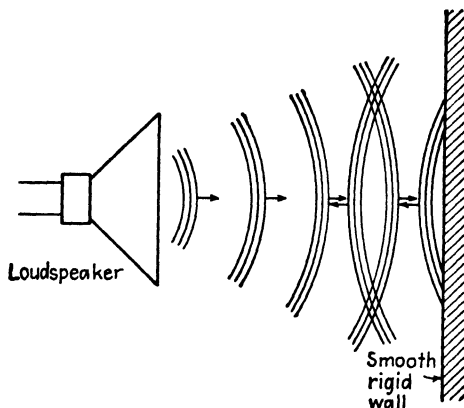



Fig. 3.—A smooth rigid wall reflects most of the sound energy striking it.

strike some discontinuity, such as a rigid wall, all the energy of the sound waves must be absorbed, or sound energy must be reflected back toward the source. If the loudspeaker of Fig. 3 is located in an enclosed room, the sound waves may continue to bounce back and forth for some time.

The length of time that the sound waves continue to exist before the sound dies out and becomes inaudible depends very much on the nature of the surface of the wall. If a rigid wall has a smooth surface (for example, it is painted wood), then about 96 per cent of the sound energy is reflected, and about 4 per cent only is absorbed by the surface.

In contrast to this, if the surface of a wall is porous and contains minute surface air pockets or interstices, then when the sound waves strike this type of wall much energy absorption occurs. This is because the air particles in the interstices are caused to vibrate, and in so doing rub against the inner surfaces of the minute air pockets. Thus they convert acoustic energy into heat energy and absorb the energy of the sound waves.

Transmission of Sound.—Very little sound is transmitted through a massive solid wall because the feebly vibrating air particles cannot impart much energy to a massive wall. If, however, the wall contains cracks or other small openings, sound energy will leak through.

If a wall or partition is of flimsy panel construction, the panels will vibrate readily, and will radiate much sound into adjacent rooms. But when a wall or a surface vibrates, the particles and fibers of material "slide along each other," and this internal friction dissipates some acoustic energy.

Confusion often occurs regarding the acoustic action of a dead-air space. An enclosed space of this kind, where the air is unable to contain large air currents, is an effective *heat* insulator, but it is a *very poor sound* insulator. Sound waves readily pass through air even if the air is enclosed.

FIG 4.—When a sound wave, here represented as a ray, strikes a rigid wall, some of the energy is reflected, some is absorbed, and the remainder is transmitted through the wall. The path of the ray through the wall is bent slightly because of refraction.

Sound waves pass through most glass windows with relative ease because a glass window pane vibrates, particularly at certain frequencies at which the pane is mechanically resonant. If windows must be placed between two rooms (for instance, a radio control room and a radio studio), the transmission of sound can be minimized by using several glass panes of different thicknesses. These panes will have different resonant frequencies of vibration, and since usually no two thicknesses will vibrate appreciably at the same frequency, the transmission of sound is well blocked. The air spaces between the panes contribute little or nothing to the stopping of the sound waves.

The phenomena occurring when a sound wave strikes a rigid wall are depicted in Fig. 4. Most of the energy of the oncoming wave (shown here as a ray of sound) is reflected by a smooth rigid wall. On the other hand, much is absorbed if the *surface* is porous. (If the entire wall is porous, then sound would be transmitted through the small openings, just as it would pass through cracks.)

If a solid wall is struck, as with a hammer, these *directly imparted mechanical vibrations* are transmitted readily by the wall. If a steam pipe in the basement is tapped with a hammer, these vibra-

tions are transmitted readily throughout a building. Machines impart vibratory energy to a building in a similar way. These are mechanical vibrations, and they may produce air-borne sound waves. The remedy for this common evil is to isolate machines from floors, or other parts of a building, with isolating mountings of rubber, cork, or similar material. In doing this, care should be taken to ensure that no part of the machine or mounting makes a solid mechanical bond (possibly through foundation bolts) between the machine and the building; that is, a machine must be *completely* isolated.

Reverberation.—If a person is outdoors and very close to a large reflecting surface (for instance, the side of a large building), sound waves will be reflected back when he speaks (as shown in Fig. 3). If the distance is less than a critical value, the person speaking will not detect the reflected wave because insufficient time has elapsed between the original and reflected wave, and the two separate waves are not recognized by the ear.

If the person now steps back from the wall and speaks, an echo will be detected if the distance exceeds a critical value. The time interval required is about 1/17 second.

An **echo** is a distinctly recognizable reflection of sound, such a a reflected spoken word. Echoes usually are noticeable outdoors where there is only one, or at most very few, reflecting surfaces. An echo can occur in a large room, but usually there are so many reflecting surfaces present that the multiple reflections cause an unrecognizable "jumble" of sounds, rather than a distinct echo. If a room does have a distinct echo, it can be reduced in intensity by covering the offending reflecting surface with some material that will absorb the sound (page 7).

If the walls, floor, and ceiling of a room consist of smooth reflecting surfaces, the absorption of sound energy will be low, the reflection will be high, and the sound waves will bounce back and forth many times. Only after a considerable period will the acoustic energy be dissipated to the point at which the waves are insufficient in strength to actuate the ear and cause sounds to be heard. This prolonging of sounds, after a sound source (such as a speaker) has ceased to emit sound, is termed **reverberation.** Excessive reverberation is bothersome because the prolonged sounds interfere with the next words spoken, causing the speech to be indistinct and difficult to understand. This occurs in poorly designed auditoriums.

If the reverberation time of a room is long, it is because the enclosing surfaces do not absorb the sound with sufficient rapidity. The only satisfactory remedy is to add such sound-absorbing material as carpets, draperies, upholstered furniture, clothing, and special acoustic wall coverings. These materials contain small openings, or interstices, that dissipate sound energy (as explained on page 4).

Photograph of four types of special acoustic wall covering used for reducing reverberation time in auditoriums and radio studios Sound energy is dissipated in the porous surfaces.

It is safe to state that all large rooms will have an excessive reverberation time unless care is taken in their design and construction. It is usually necessary to add sound-absorbing wall coverings to keep the reverberation time sufficiently low so that music will be pleasing and speech will be clear and understandable. All radio studios must be acoustically treated. Rooms often are said to have bad echoes when in reality the trouble is excessive reverberation time.

Calculation of Reverberation Time.—It is possible to predetermine the reverberation time of a room, such as a radio studio or an auditorium. Prior to construction, calculations can be made that will show if the reverberation time will be correct.

The reverberation time of a room, *such as an auditorium*, can be calculated by the formula developed by W. C. Sabine

$$T = \frac{0.05V}{a}, \tag{1}$$

where T is the reverberation time in seconds, when V is the room volume in cubic feet, and a is the total sound-absorbing ability of *all* surfaces in the room that are exposed to the sound waves. This equation applies with accuracy only to rooms and auditoriums that have a reverberation time of several seconds.

Most radio studios, particularly small ones, should have rever-

beration times of about 1 second or less. Therefore Eq. (1) does not give the correct value for studios. For such studios, a value of $0.027 V/S$ should be subtracted from Eq. (1). In this, V is the volume in cubic feet, and S is the total room surface in square feet.

By definition, the **reverberation time** is the length of time required for a sound to decrease in intensity to one-millionth (60 decibels, page 95) of its original value. From a practical standpoint, the reverberation time is the time required for a very strong sound to decrease to the point at which it is inaudible. Thus the reverberation time of a room can be measured approximately by blowing a horn loudly and checking with a stop watch the length of time the sound remains audible after the blowing is stopped suddenly. This method is difficult to apply to a radio studio that has a reverberation time of about 1 second or less, because of the short time interval. It is a good practical check on reverberation, however.

Because the sound-absorbing abilities of room surfaces vary with frequency, the reverberation time also will vary with frequency. Calculations and studies sometimes are made at frequencies (selected on the musical scale) of 128, 256, 512, and 2048 cycles. Often they are made only at 128 and 512 cycles, and sometimes only at 512 cycles. Most auditoriums should have special sound-absorbing materials added to the interior surfaces to reduce the reverberation time to the desired value. *All radio studios must be so treated.* The amount of sound-absorbing material that must be added depends, among other things, upon the sound-absorbing ability of the material to be added.

Detailed tables of the sound-absorbing coefficients of various materials will be found in textbooks on acoustics and in communication handbooks.[1] A few values are listed in Table I. These are special sound-absorbing products made of materials that offer porous surfaces to the sound waves. Such products often are perforated, grooved, or otherwise formed so that they have greater absorbing capabilities, and so that they may be painted and yet retain much of their original sound-absorbing properties.

[1] Knudsen, V. O., "Architectural Acoustics," John Wiley & Sons, Inc.; H. Pender and K. McIlwain (eds.), "Electric Communication and Electronics," 3d ed., Vol. 5 of "Electrical Engineers' Handbook," John Wiley & Sons, Inc., 1936.

See also descriptive pamphlets published by the various manufacturers of acoustic material.

TABLE I.—SOUND-ABSORPTION COEFFICIENTS OF TYPICAL MATERIALS AND OBJECTS[1]

Brand or make	Thickness, inches	Sound-absorption coefficients at various frequencies, cycles					
		128	256	512	1024	2048	4096
Sabinite (an acoustic plaster)	0.5	0.08	0.14	0.18	0.25	0.31	0.35
Absorbex, Type A. (an acoustic tile cemented to plaster, two coats of oil paint sprayed on)	1.0	0.07	0.22	0.51	0.91	0.77	
Acousti-celotex, Type B. (a perforated acoustic tile, cemented to plaster board)	0.625	0.12	0.24	0.47	0.73	0.78	
Acousti-celotex, Type BBB. (a perforated acoustic tile cemented to plaster board)	1.25	0.19	0.41	0.91	0.92	0.92	
Fir-tex (an acoustic tile nailed to wood strips)	0.5	0.15	0.41	0.34	0.35	0.43	0.48
Glass		0.035		0.027	..	0.02	
Concrete, painted		0.009	0.011	0.014	0.016	0.017	0.018
Plaster	0.75	0.038	0.049	0.060	0.085	0.043	0.056
Wood floor	0.75	0.09	...	0.08	..	0.10	
Carpets, lined	..	0.10	...	0.25	...	0.40	
Draperies, velure, draped to ½ area	....	0.14	0.35	0.55	0.72	0.70	0.65
Chair, plywood	....		0.19	0.24	0.39	0.38	
Chair, heavily upholstered, theater	..		3.4	3.0	3.3	3.6	
Person, adult, fully dressed	..	1.8		4.2		5.5	

[1] From "Electric Communication and Electronics," Vol 5 of "Electrical Engineers' Handbook."

Recommended Reverberation Time.—As has been stated, the reverberation time of an auditorium or a studio can be calculated by the use of Eq. (1) and of Table I. The question arises: What is the recommended reverberation time of an auditorium or a radio studio? These values can be obtained from Figs. 5 and 6. These figures are the result of studies made on auditoriums and studios that have proved to be satisfactory.

Calculations of Reverberation Time.—As an illustration of the use of Eq. (1), the reverberation time will be calculated for a concert room in which radio programs are to originate.

Illustrative Problem.—A concert room is 40 feet wide, 70 feet long, and 25 feet high. It is desired to know what the reverberation time will be if the room is equipped with 200 square feet of draperies, if 75 per cent of the floor is carpeted and the rest of the floor is wood, if there are 350 theater-type chairs or seats, and if there is an audience of 250 adults. The performers will

be neglected. The purpose of this calculation is to ascertain if special sound-absorbing material must be added, and therefore it will be assumed that the ceiling and walls are covered with ordinary plaster. The trial calculation will be made at a frequency of 512 cycles, a figure often used.

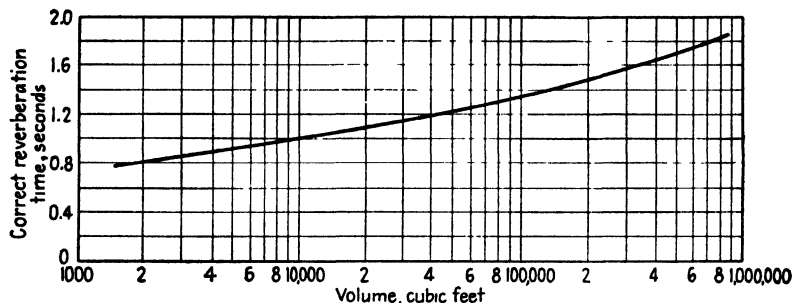



FIG. 5.—The correct reverberation time for an auditorium or music hall where radio programs are produced can be determined from this curve for frequencies of about 1000 cycles. The reverberation time for lower frequencies may well be greater than this curve indicates. If a sound-amplifying system is used in an auditorium, the reverberation time should be slightly less than shown. (*From W. A. McNair, Optimum Reverberation Time for Auditoriums, Journal of the Acoustical Society of America, Vol.* 1, *No.* 2.)

Step 1. Calculate the volume of the room and the areas of the various surfaces.

Volume = 25 × 40 × 70 = 70,000 cubic feet.
Area of ceiling = 40 × 70 = 2800 square feet.
Area of floor = 40 × 70 = 2800 square feet.
Total wall area = 25 × 220 = 5500 square feet.
Area of draperies = 200 square feet.

Step. 2. Compute the total amount of sound absorption present. To do this, the various areas will be multiplied by the coefficients at 512 cycles as given in Table I.

Floor, 2800 × 0.75 × 0.25	=	525 units
2800 × 0.25 × 0.08	=	56 units
Ceiling, 2800 × 0.06	=	168 units
Walls, 5500 × 0.06	=	330 units
Draperies, 200 × 0.55	=	110 units
Chairs, 100 × 3.0	=	300 units
Audience, 250 × 4.2	=	1050 units
Total..................		2539 units

In making these calculations, only 100 theater chairs were figured in because it was assumed that the 250 persons present were seated and thus the occupied chairs offered little absorption. Also, windows (if any)

were neglected because the absorption of glass is not sufficiently different from plaster to affect the results in this case.

Step 3. Use Eq. (1) and the correction discussed on page 8 to calculate the reverberation time.

$$T = \frac{0.05V}{a} - \frac{0.027V}{S} = \frac{0.05 \times 70{,}000}{2539} - \frac{0.027 \times 70{,}000}{11{,}100} = 1.2 \text{ seconds.}$$

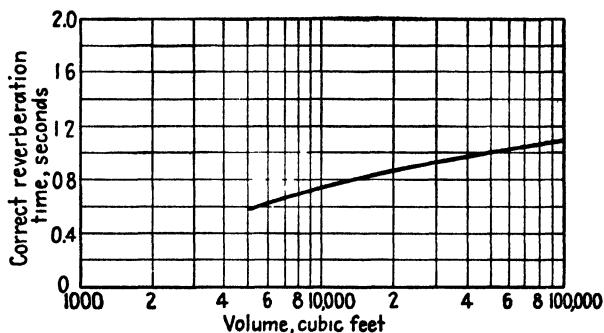



FIG 6.—Correct reverberation time for radio studios. This curve applies in the vicinity of 1000 cycles (*Data from Acoustics of Radio Broadcasting Studios, Journal of American Institute of Electrical Engineers, Vol.* 49, *No.* 3.)

An examination of Fig. 6 indicates that the reverberation time for a studio of 70,000 cubic feet should be about 1.0 second for good results. Thus at first thought it may seem that the studio would be fairly satisfactory.

Before such a conclusion is drawn, however, conditions should be investigated when no audience is present, because this situation undoubtedly will be encountered. When no audience is present, 1050 units would be subtracted but 250 × 3.0 = 750 units would be added (250 more chairs would be exposed), leaving a net loss of 300 units. For this condition, the reverberation time would be

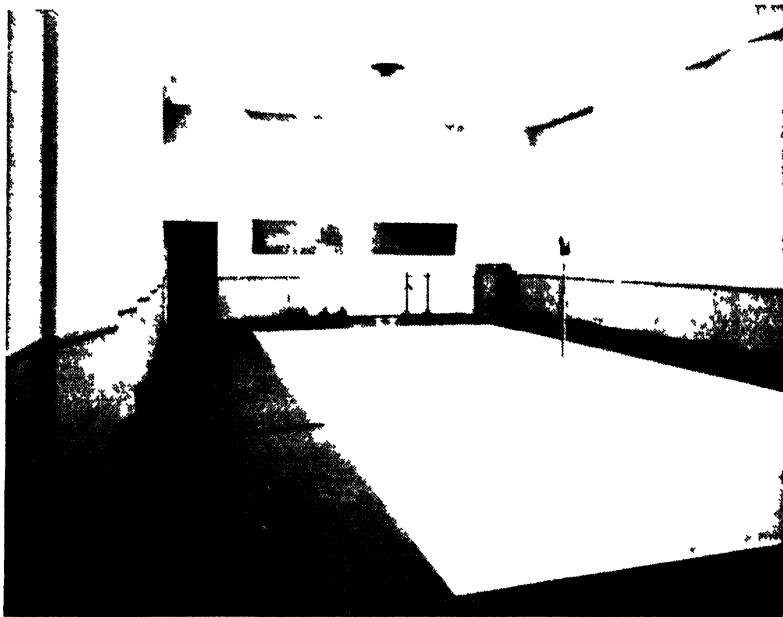
$$T' = \frac{0.05 \times 70{,}000}{2239} - \frac{0.027 \times 70{,}000}{11{,}100} = 1.4 \text{ seconds.}$$

The two calculations show the advantage of using upholstered chairs to provide much of the sound absorption. When a studio is empty, the chairs are effective and keep the reverberation time low. When the audience is present, most of the chairs are covered and the audience provides much of the sound absorption. Thus, to a large extent, the reverberation time in this particular room is independent of the audience.

It is decided to reduce the reverberation time to 0.9 second with an audience of 250. To obtain this reverberation time will require

the addition of sound-absorbing material. The number of units that must be added can be found from the calculation

$$0.9 = \frac{0.05 \times 70{,}000}{2539 + x} - \frac{0.027 \times 70{,}000}{11{,}100} \quad \text{and} \quad x = 733 \text{ units.}$$

From a practical standpoint, this sound-absorbing material can be added anywhere, provided that it is exposed to the sound

A large radio studio It will be noted that the walls are broken up into large splayed panels so that the parallel surface area in the room is reduced These irregular surfaces diffuse the sounds and reduce flutter (*National Broadcasting Company*)

waves. A study of the situation shows that it is convenient to add to the rear wall 600 square feet of material which has an absorption coefficient of approximately 0.5, giving a total absorption of 300 units, and also that it is convenient to add to the side walls panels of sound-absorbing material which have a total area of 1500 square feet and an absorption coefficient of 0.3, giving a total absorption of 450 units. These two treatments combine to give 750 units, a value slightly greater than that needed. A further investigation might indicate the desirability of placing some

additional sound-absorbing material on the ceiling and of using lighter draperies and possibly a lighter rug. Sometimes parallel surfaces cause a back-and-forth reflection, which produces a "fluttering" effect. This can be prevented by not using large parallel surfaces, or by covering such surfaces with sound-absorbing material if they must exist.

Music needs a certain amount of reflection and reverberation to blend the tones and "mellow" the sounds properly. Without this, music sounds harsh. For this reason, the so-called "live-end, dead-end" studio sometimes is used. With this arrangement but little acoustical treatment is used at one end, where the musicians perform, and much sound-absorbing material is used at the other end, where the audience is seated or where the microphones are placed. Very small studios for speech and announcements hardly can be made too dead because no extensive blending is required.

Sounds should not be transmitted to a studio from outside. The use of several layers of glass for windows was mentioned previously. It is common practice to "float" studios on suitable rubber or cork mountings so that they are not in direct contact with the rest of the building. Such mountings largely prevent the transmission of building vibrations to the studio. Because studio windows, if they exist, are usually sealed it is necessary to provide forced ventilation. This should be done through air ducts that are lined with sound-absorbing material, and low air velocities should be used. Much care must be taken in the design of radio studios to ensure that noise does not originate within the studio and that it is not transmitted in from outside.

Noise.—A radio studio *must* be free from noise. The floors should be covered, so that noise will not be produced to any marked degree. Furniture and equipment should be of types that produce little noise. If these precautions are taken, the sound-absorbing material present in a studio quickly will absorb and damp out the unavoidable noises that are produced.

Noise is defined as any unwanted sound. Thus music may be noise if it is unwanted. The noise level at a given location is measured with a **noise meter.** Since this device may be used also to measure any sound, it is usually called a **sound-level meter.**

A block diagram of a sound-level meter used for measuring noise is shown in Fig. 7. The microphone picks up the noise, or sound, to be measured and impresses a corresponding electric signal on

the preamplifier. The variable attenuator alters the sensitivity of the device so that noise levels of greatly differing magnitudes can be measured. The frequency-weighting network gives the sound-level meter characteristics that are somewhat like those of the human ear. As will be seen later (page 18), the ear is not equally sensitive to all frequencies, or at all sound levels. The ear is insensitive to both low frequencies and high frequencies at ordinary sound intensities. If a weighting network were not introduced, a sound-level meter that had uniform response might measure a low-frequency noise or a high-frequency noise as quite

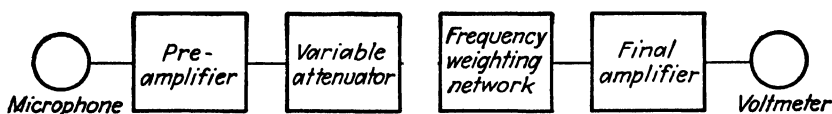



FIG. 7.—Schematic diagram of a sound-level or noise meter. The weighting network gives the device characteristics approximating those of the human ear.

intense, but the ear, because of its nonuniform response, would judge the same noise as being of low intensity. A final amplifier increases the weighted signal and impresses it on a voltmeter, often of the vacuum-tube type.

Noise and sound levels are measured in decibels (page 97) above a reference intensity, or zero level, of 10^{-16} watt per square centimeter. This value is approximately the magnitude of the weakest sound the ear can hear under ideal conditions. Thus, from a practical viewpoint, zero on a sound-level meter represents the threshold of hearing. A sound-level meter often is used in radio contests to measure applause.

As previously mentioned, adequate sound-absorbing material rapidly damps out noise. Thus if an office, reception room, or similar location is noisy, the addition of sound-absorbing material will reduce the noise level. If the total amount of sound absorbing material (as calculated on page 10) is a_1 units, then the addition of sound-absorbing material giving a final total of a_2 units will reduce the noise level by the amount

$$\text{Reduction in decibels} = 10 \log_{10} \frac{a_2}{a_1}. \tag{2}$$

A reduction of from about 5 to 10 decibels by the addition of sound-absorbing material is about the limit usually attempted. Further reduction should be achieved by quieting sources of noise.

Speech Sounds.—The wave form of a pure tone produced by a wire vibrating at the fundamental frequency was shown in Fig. 2. Speech sounds (and those of music as well, as will be discussed later) are of course very much more complicated than the simple sound shown in Fig. 2. Figure 8 is included as an example.

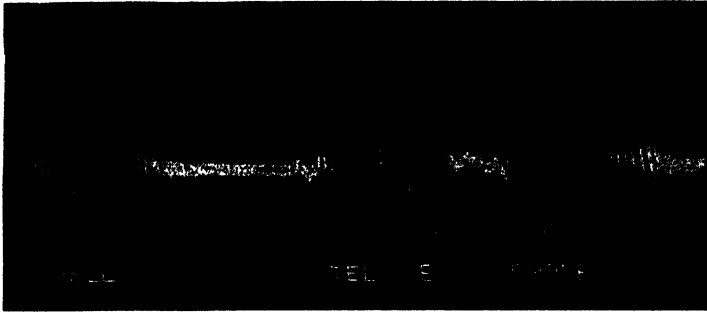



Fig. 8.—Speech and music sounds are far from being pure sine waves. Typical speech sounds have the wave form indicated. (*The Bell Telephone System.*)

The origin of the sounds in speaking and singing is as follows: (*a*) those sounds produced by the vibrating vocal cords and modified by the resonating air cavities of the head, (*b*) those produced by passing air through small openings or over sharp edges in the mouth (for example, the teeth), and (*c*) sounds produced by a

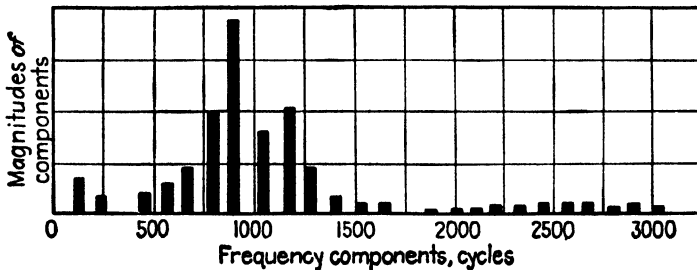



Fig. 9.—Speech and music sounds are composed of a fundamental and various harmonics or overtones. This illustration shows the relative magnitudes of the various frequency components present in the letter a, pronounced as in tär. (*Courtesy of D. Van Nostrand Company, Inc.*)

combination of (*a*) and (*b*). The vibrating vocal cords modulate the air flowing from the lungs and produce basic tones that are rich in harmonics, or overtones (see Fig. 9). These tones are suppressed or enhanced by the various cavities in the throat, mouth, and nasal passages. Vowel sounds are produced in this

way. The other sounds are produced largely by methods (*b*) and (*c*).

As Fig. 8 indicates, the frequencies and amplitudes of speech sounds vary widely. The *average speech power* in normal talking is about 10 microwatts. In speaking and singing, the values fall below, and rise above, this value. A whisper is of the order of

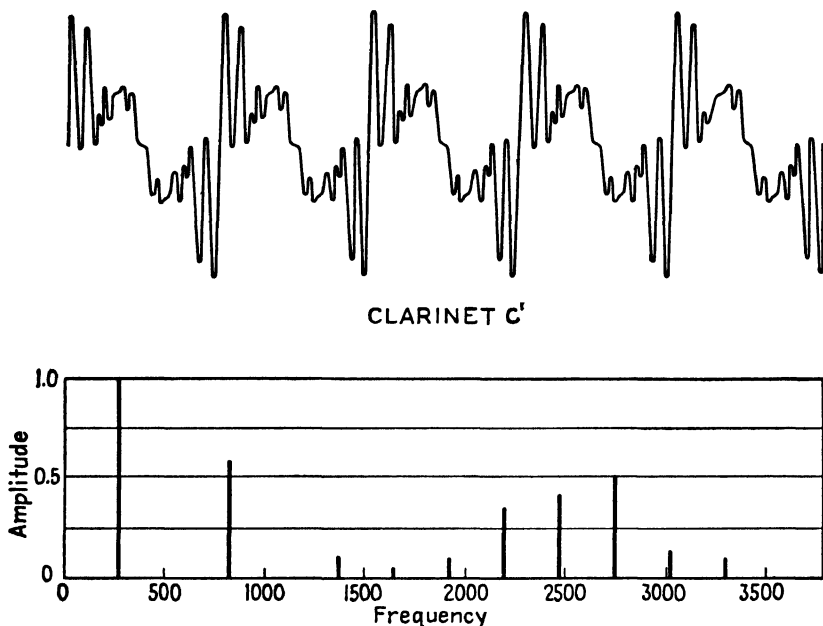



Fig. 10.—Wave-form and frequency analysis of the musical note C' when played on a clarinet. (*Bell Telephone Laboratories Inc., and D. Van Nostrand Company, Inc.*)

0.1 microwatt, and peak values of 1000 microwatts or more occur, depending of course on the individual. The ratio of the strongest to the weakest speech sounds is about 40 decibels (page 95), which is a power ratio of 10,000 to 1. The frequencies found when speaking or singing depend on the individual, but most of these sounds are contained within a band of from 60 to 10,000 cycles.

Music.—A vibrating wire, as in a piano, was depicted in Fig. 1 as producing a pure sound that has the sinusoidal wave form of Fig. 2. In general, taut wires and strings do not vibrate with the simple motion of Fig. 1. Musical instruments of all types are so designed and operated that **harmonics, or overtones,** are produced along with the fundamental. Harmonics are components having

frequencies of $2f$, $3f$, $4f$, etc., where f is the frequency of the fundamental. On page 1, sound waves were shown to have the distinguishing characteristics of frequency (pitch) and intensity; that is, sounds differ from each other in these respects. A third distinguishing characteristic is **quality.** The quality of a sound is determined largely by the number and magnitude of the harmonics

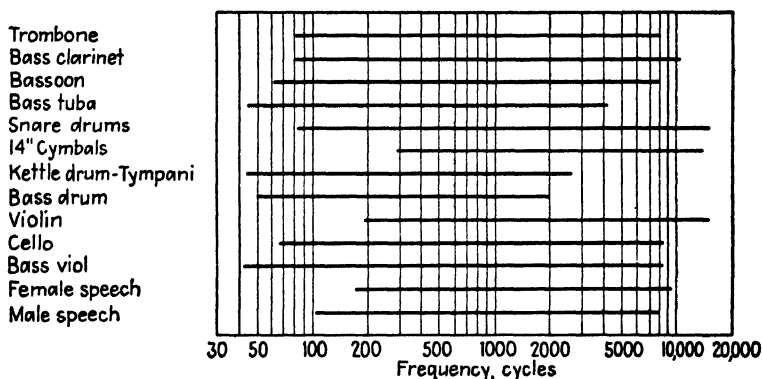



FIG. 11.—Frequency ranges of certain musical instruments and of speech. (*Data from Long Distance Cable Circuit for Program Transmission* [*A. B. Clark and C. W. Green*], *Transactions of the American Institute of Electrical Engineers, Vol.* 49, *No.* 8, *August*, 1930.)

(overtones) and by their frequency distribution. An example of this is shown in Fig. 10.

For most musical instruments the number of harmonics, or overtones, is very great. This is particularly true of string instruments, such as the piano and violin. Also, the player can vary the basic frequency (or tone) over a wide range, producing basic tones of very high frequency, which are accompanied by their overtones. Furthermore, such devices as long organ pipes, etc., produce very low tones. The result is that the sounds produced by musical instruments cover a wider range than those produced by the voice, even in singing.

Just how low and how high in frequency do musical sounds go? This point can be argued, because if measuring instruments of sufficient sensitivity were employed, components of extremely high frequency would be found. But from a practical standpoint, it can be said that most of the sounds of music lie in a band between 40 and 15,000 cycles, as Fig. 11 indicates. As will be seen in the next section, a band of this width is not entirely necessary for

excellent reproduction of music. The sound level in the rendition of a musical program may be very high compared with speech, particularly if a large orchestra or a band is playing. Also, the variations in volume may be as great as 50 decibels or more for an orchestra. This corresponds to a change in intensity of 100,000 to 1 (page 95).

Radio circuits need not be designed to transmit bands of frequencies to include all components; nor do they need to handle such extreme volume changes. A discussion of these matters involves a consideration of the phenomenon of hearing, and this will be treated in the following pages.

Hearing.—The human ear is, in a sense, the terminating element of the usual radio channel. The ear consists of three major parts:

The outer ear has an external portion for diverting sound into the auditory canal leading to the eardrum at the end. The eardrum is a thin membrane that is caused to vibrate by the sound waves striking it.

The middle ear contains a mechanical lever system (hammer, anvil, and stirrup) for decreasing the relatively large vibrations of the eardrum and transmitting these reduced vibrations to the inner ear.

The inner ear consists of several parts, one of which is the cochlea, a canal system filled with fluid. The fluid of the inner ear is set in vibration by the mechanical system of the middle ear, and apparently transmits the impulses to various nerve endings that terminate in the fluid. Impulses of certain frequencies seem to excite certain nerve endings only, and in this way the various frequencies are recognized. The nerves mentioned lead to the brain.

In the design and operation of a radio system, attention is focused on the response characteristics of the ear, that is, on the sensitivity and frequency response and on the effects of noise.

The sensitivity of the ear is remarkable. The power level of sounds that are just audible in an absolutely quiet location is very low. From a practical standpoint, the **threshold of audibility** (that is, the power level at which a sound is first audible) for the average normal ear is about 10^{-16} watt per square centimeter at 1000 cycles, in the absence of all interfering sounds.

The frequency response of the average normal ear is from about 20 to 30,000 cycles for some youths, and it is from about 20 to 20,000 cycles for the average adult. With increasing age the

range decreases, particularly the upper limit. Partial deafness caused by illness or injury may reduce the frequency response considerably. Many persons cannot hear above a few thousand cycles and do not suspect the fact.

The judgment of loudness, that is, the magnitude of the sensation produced in the brain, is of interest. A sound that seems quite loud to one person may sound very weak to a person who has impaired hearing.

The masking effect of sounds is very important in a radio system transmitting information. Noise (which was defined as any unwanted sound) tends to mask out the desired sounds, and may reduce their understandability. Such action interferes greatly with the transmission of information, and it may render a radio channel inoperative. Interfering noises may be acoustical in nature, or may be from electrical sources, for instance, hum, static, etc.

Fidelity of Performance.—In the preceding pages, it was explained (*a*) that sounds produced by the human voice may vary in intensity in the ratio of 10,000 to 1 and that they cover a frequency range of at least 60 to 10,000 cycles, and (*b*) that in music the intensity variation was as great as 100,000 to 1 and that the frequency range was from at least 40 to 15,000 cycles. Of course in some instances there are components below, and particularly above, these values.

The question immediately arises: What must be the fidelity of a radio system? By the **fidelity** of a radio system is meant the degree with which a system (or a portion of a system) accurately reproduces at its output the original input signal. The necessary degree of fidelity is a debatable question, and one that has not been settled. Several studies have been made on this subject, and these will be considered now.

Speech.—Articulation tests have been made on speech by transmitting and listening to sounds *not conveying ideas.* The percentages of the sounds correctly recorded by observers is a measure of the **articulation.** As a result of such tests it was found that 86 per cent of the sounds were understood if all frequencies below 1000 cycles per second were missing and that 96 per cent were understood if all frequencies below 500 cycles were omitted. This shows conclusively that low-frequency components, below about 500 cycles per second, are not necessary for the transmission of

intelligible speech. Similar tests showed that the high-frequency components, above about 4000 cycles per second, were not necessary for good intelligibility. From the standpoint of transmitting information only, a band of frequencies of from about 200 to 4000 cycles gives good intelligibility.

Music.—Two important tests have been made on music to determine the fidelity characteristics a system should have. A study was made, about 1930, to determine how much the band of transmitted frequencies could be narrowed *before impairment of quality was observed.* The required volume range also was studied at this time. A second study, reported in 1945, investigated the width of transmitted frequency band *that was preferred* by listeners. These two studies will be reviewed now.

Impairment of quality was studied[1] by having observers listen to high-quality speech and music from different instruments, when the music was reproduced with substantially the entire frequency range transmitted, and when either high frequencies or low frequencies were eliminated by electrical filters. The observers noted when the elimination of low frequencies or high frequencies affected *the quality of the sounds* heard. They did not attempt to determine if the sounds heard were more or less pleasing, but merely noted when an impairment of quality was apparent. These tests showed that little quality was lost if a frequency band of from 50 to 8000 cycles was transmitted. This same article reported that a transmitted volume range of 40 decibels (power ratio of 10,000 to 1) was satisfactory.

Preferred frequency range was studied[2] by having the listeners judge broadcast-program material, both speech and music, and indicate *the frequency band width preferred.* The reproduced radio program material to which the observers listened was of three band widths: *wide range,* from about 30 to 10,000 cycles; *medium range,* from about 60 to 6000 cycles; and *narrow range,* from about 120 to 4000 cycles. As a result of this study, which involved

[1] Reported in an article by W. B. Snow, Audible Frequency Ranges for Music, Speech, and Noise, *Journal of the Acoustical Society of America,* Vol. 3, July, 1931. See also A. B. Clark and C. W. Green, Long Distance Cable Circuit for Program Transmission, *Transactions of the American Institute of Electrical Engineers,* Vol. 49, No. 8, August, 1930.

[2] Reported in an article by H. A. Chinn and P. Eisenberg, Tonal Range and Sound-intensity Preferences of Broadcast Listeners, *Proceedings of the Institute of Radio Engineers,* Vol. 33, No. 9, September, 1945.

almost 500 listeners, *it was found that the listeners preferred either the narrow or medium tonal range to a wide one.* Also, it was found that the listeners still preferred a narrow tonal range to a high tonal range, even when informed that one condition was low fidelity and that the other one was high fidelity. Related tests showed that the listeners preferred a sound-intensity level somewhere between 60 and 70 decibels above zero reference level (page 97). For comparison, 60 decibels would be about the level of loud *conversational* speech. It should be mentioned that some authorities do not accept as final the results of these tests.

The matter of fidelity of performance has been considered here in some detail because it is of great importance. The general opinion in radio seems to be that a wide range of frequencies must be transmitted and that the radio listeners prefer their programs with wide-band transmission. The tests reported seem to show that (*a*) if a band of from about 50 to 8000 cycles is transmitted, the average listener would not be able to detect the fact that certain low-frequency, and high-frequency, components were missing, and (*b*) that the average listener prefers programs of narrow or medium tonal range.

Microphones.—The microphone is an electroacoustical device which is actuated by sound waves and which develops electric signals corresponding to the sound waves. There are two basic types of microphones, the modifier types and the generator types.

Modifier microphones are actuated by the acoustic energy of the sound waves, and they *control* the flow of electric energy from a source such as a battery. The sound waves *modify* the current in the circuit and cause the current to vary as the sound waves vary. The electric signals at the output of a modifier microphone are replicas of the sound waves. The output of a modifier microphone is relatively high because the sound waves *control* energy flow from a separate battery source. The carbon-granule microphone is an example of this type.

Generator microphones are actuated by the energy in the sound waves, and *no* additional source of energy (such as the battery used with carbon-granule microphones) is connected in the microphone circuit. The generator-type microphone is driven by the acoustic energy of the sound waves, and it generates an electric signal that corresponds to the sound waves. Since sound waves contain but little energy, and since the efficiency of this microphone

in converting from acoustic to electric energy is low, the generator-type microphone has a very low electric output. It was shown on page 16 that average conversational speech was at a level of about 10 microwatts; it follows, from the preceding discussion, that the electric-power output of a microphone would be considerably below this when actuated by speech of normal intensity.

Carbon-granule Microphones.—These were the first microphones used in radio systems, and were once used extensively in radio, including broadcasting. At present, their radio use is restricted to a large extent to communication systems for the

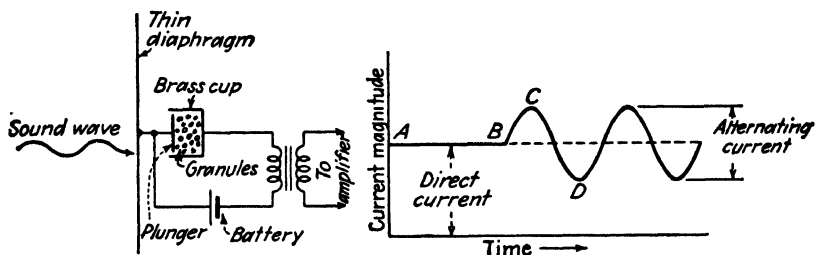



Fig. 12.—When sound waves strike the diaphragm of the single-button carbon-granule microphone, the diaphragm and the attached plunger move back and forth. This action varies the resistance of the carbon granules and causes the current from the battery to vary in accordance with the sound waves.

transmission of information by speech, as in some airways radio-communication systems. Carbon-granule microphones are of two types, single-button microphones and double-button microphones.

Single-button carbon-granule microphones are of the type used in telephone systems, and sometimes used in radio if information only is to be transmitted and if intelligibility is of importance, but high quality is not necessary. In its simplest form, a single-button carbon-granule microphone is arranged as in Fig. 12. The battery often is a few dry cells, which gives 3 to 6 volts, and the current is of the order of 0.1 ampere.

A thin diaphragm is attached to an electrode that moves in accordance with vibrations of the diaphragm. This electrode acts as a plunger in compressing the carbon granules in a brass cup. The cup is sealed with a flexible ring so that the granules will not fall out, and it is often called a "carbon button." In modern telephone microphones the diaphragm also acts as the front electrode.

When no sound waves strike the diaphragm, the mass of carbon

granules is at rest, and the resistance offered to the electric current forced through the circuit by the battery is substantially constant. This current is shown from A to B of Fig. 12. When sound waves strike the diaphragm, as shown in the figure, they cause the diaphragm and the attached electrode to move in and out. The inward motion increases the pressure on the granules, and this *decreases* their resistance, causing the current to rise as indicated

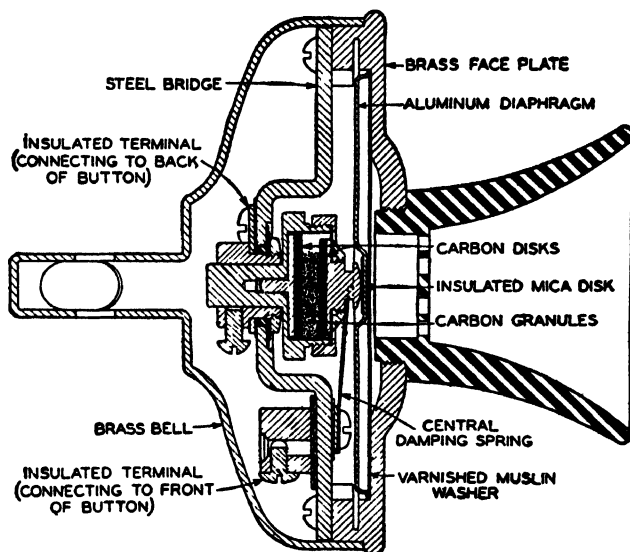



FIG. 13.—Cross section of a single-button carbon-granule microphone of the older type.

at point C. The outward motion of the diaphragm and the attached electrode reduces the pressure on the granules, and this *increases* their resistance, causing the current to fall as at point D of Fig. 12.

In this way, sound waves control the current flow from the battery and produce current variations that correspond to the sound waves. These current variations induce a corresponding signal voltage in the transformer secondary, and this voltage can be amplified by vacuum tubes and then used to control the output of a radio transmitter.

Although considerable distortion exists when a single-button microphone is used, it is satisfactory where speech only is to be transmitted. The main advantages of this microphone are its

simplicity, ruggedness, and particularly its high output. In certain low-powered radio transmitters very little speech amplification need be used between the microphone and the radio-frequency transmitting portion.

If the microphone of Fig. 12 is held in a position so that the diaphragm is horizontal, it is possible that the granules will partly, or entirely, fall away from the top electrode, thus greatly reducing or entirely interrupting the current flow. Thus the operation of such a device would depend greatly upon the position in which it was used. Microphones can be made essentially nonpositional by constructing the front electrode and the carbon-granule chamber so that the current path largely is independent of the position of the microphone. Typical constructions of the old and new microphones are shown in Figs. 13 and 14.

FIG. 14.—Cross section of a carbon-granule microphone used in modern telephone handsets. The construction of the carbon-granule chamber is such that the granules cannot fall entirely away from the electrodes and open the circuit. This feature makes the microphone nonpositional.

Double-button carbon-granule microphones are used where the quality of the transmission must be higher than that given by the single-button type. Two buttons, one on each side of the diaphragm, are used, and they are connected in a circuit, as shown in Fig. 15. Several dry cells may be used in series to supply the microphone current, which is a total of 0.025 ampere for a typical double-button microphone.

When no sound strikes the diaphragm, a constant current flows through each half of the transformer primary. When sound waves move the diaphragm in to the right, the pressure on the granules in the cup on the right increases, and the current I increases through this cup and the upper half of the transformer primary. At the same instant the pressure on the left cup decreases, and the current I' decreases through this cup and the lower half of the transformer. When current I flowing in *one* direction increases in one half of the primary, and current I' flowing in the *other* direction decreases in the other half of the primary, the magnetic effects

of each primary current are additive; and the voltage induced in the secondary is twice that produced by one-half of the primary. This produces, say, the positive half cycle. When the sound waves cause the diaphragm to move out to the left, the action previously explained is reversed, and the negative half cycle is induced in the secondary.

The action just explained somewhat resembles that of the push-pull amplifier (page 333). This push-pull action reduces distortion caused by even-numbered harmonics, such as the second harmonic. Also, the double-button microphone reduces distortion in another

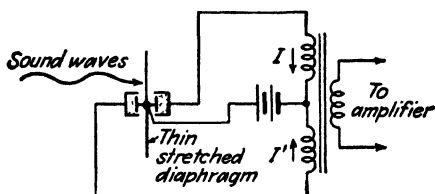



FIG. 15.—In the double-button carbon-granule microphone the distortion is reduced because there are two buttons, and the diaphragm motion is more uniform. Even harmonics (such as the second) are canceled in the transformer primary.

way. To refer to the *single-button* microphone for a moment, when the diaphragm moves in, it has to compress granules, but when it moves out it does not compress granules. As a result, when a uniform sound wave actuates a single-button microphone, unequal travel results, and as a result one half cycle of the created electric current wave will be different from the other. But with the double-button type the motions in the two directions are the same, because for *each* direction of travel the granules of one button are being expanded and those in the other are being compressed. Thus the current half cycles are quite closely identical.

A cross-sectional view of a double-button carbon-granule microphone that was widely used in radio broadcasting and sound-amplifying systems is shown in Fig. 16. The electric output of this microphone is very high compared with generator-type microphones that are used to a large extent in radio today. The output, however, is not so high as that of the single-button telephone type. This is just the opposite from what might be expected, and the reason for this is the following: In general, single-button microphones (such as those used in wire telephony for speech only) are designed for high output, and *quality* is sacrificed for high *quantity*. The double-button microphone, such as the one shown in Fig. 16,

is, on the other hand, a precision device, and in its design high *quantity* of output was sacrificed for high *quality*. In fact, a microphone having very high quality was produced by using a thin, tightly stretched diaphragm which had a high natural frequency of vibration and by damping the diaphragm motion so that it would not vibrate excessively at any frequency.

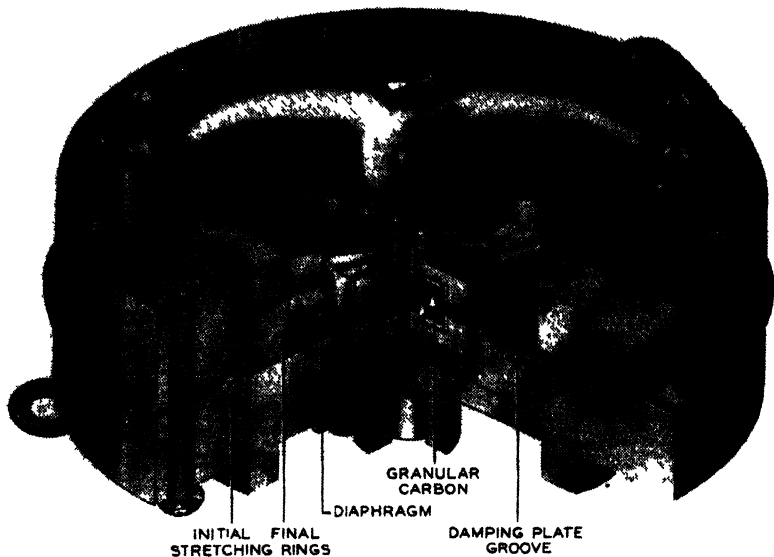



FIG. 16.—Cross-sectional view of a double-button carbon-granule microphone. (*Courtesy of Bell System Technical Journal*.)

If this is true, two questions arise: Why are double-button carbon-granule microphones not used today in radio broadcasting systems, and why are such microphones largely used for speech only? One answer is that carbon microphones have a background carbon "hiss," or noise, that is objectionable. The path of the current through the granules is not exactly constant, but varies erratically. This causes small changes to occur in the resistance of the path, even when the microphone is not being actuated by sound waves, and, hence, the microphone produces a hissing sound in the loudspeaker, which is particularly annoying when listening to music, but which is of little importance when listening to the voice transmission of information. The output characteristic curve of the microphone of Fig. 16 is shown in Fig. 17.

Zero level for this curve is one bar (or barye), which is a unit used to measure air pressure and the intensity of sound waves. One bar is equal to one dyne per square centimeter. The average intensity of *speech* sound waves is about 0.4 bar. For music the range is from about 0.5 to 1250 bars.[1]

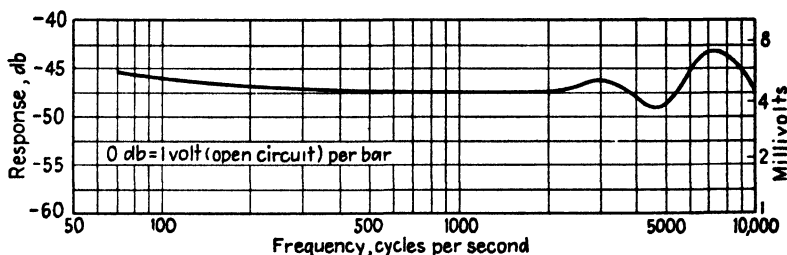



FIG. 17.—Frequency characteristics of the double-button carbon-granule microphone of Fig. 16. For the meaning of the term "bar" see above. The output of this microphone is high, and its quality is good, but the "carbon hiss" is objectionable for some applications.

Condenser Microphones.—Early in the history of radio, these microphones replaced the carbon-granule type for broadcasting purposes. The condenser microphone is similar to the carbon-granule type in that a source of direct voltage (of about 200 volts) is placed in series, as shown in Fig. 18. The condenser microphone

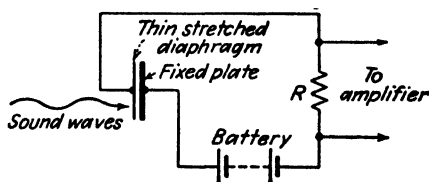



FIG. 18.—In the condenser microphone the sound waves striking the diaphragm cause the capacitance between the diaphragm and the fixed plate to vary. This varying capacitance causes corresponding charging currents to flow. The resulting IR drop across resistor R is a voltage corresponding to the sound waves.

differs fundamentally from the carbon-granule type in that a direct current does not flow through it.

The condenser microphone consists essentially of a thin metallic diaphragm spaced about 0.001 inch from a fixed metal plate, and insulated from it. Sound waves striking the diaphragm cause it

[1] H. Pender and K. McIlwain (eds.), "Electric Communications and Electronics," 3d ed., Vol. 5 of "Electrical Engineers' Handbook," John Wiley & Sons, Inc., 1936.

to vibrate, thus changing the capacitance between the diaphragm and fixed plate. When the capacitance changes in accordance with the sound waves, the electric charges on the plate must vary, because of the fact that $Q = CE$; that is, the quantity Q of electricity on the plates of a condenser is the product of the capacitance C and of the voltage E. As the quantity varies, the "charging

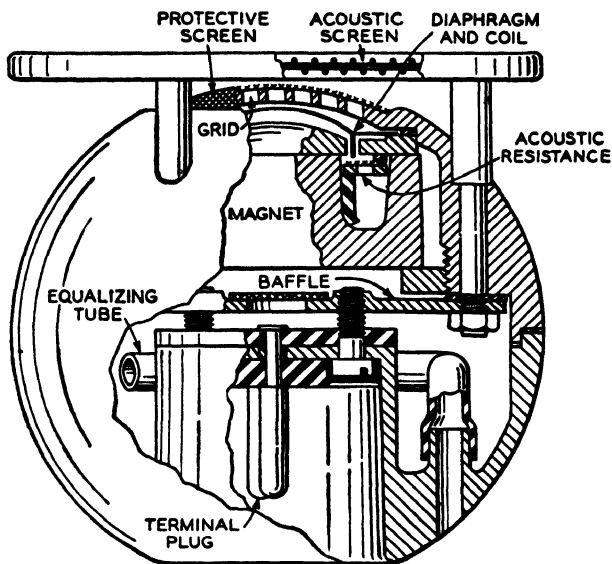



FIG. 19.—Cross-sectional view of a dynamic microphone. Sound waves striking the diaphragm cause it to vibrate, thus moving a small coil in a strong magnetic field. By this action a voltage corresponding to the sound waves is induced in the coil. This particular microphone is placed in a spherical housing to reduce distortion of the sound waves. The diaphragm is horizontal so that it will respond equally well to sounds from any direction in the plane of the diaphragm. (*Bell Telephone Laboratories, Inc.*)

current" flowing through the high resistance R (often many million ohms) varies. This current causes an alternating-voltage drop across resistance R, and this voltage drop is a close replica of the sound waves striking the diaphragm. Thus sound-wave variations are changed to electric-voltage variations, and this signal voltage can then be amplified and used to control a radio transmitter. Usually the amplifier is located very close to the microphone.

The quality of the condenser microphone is excellent, and its operation is quiet as compared with the carbon-granule type. Its output level is very low. Although for some years the condenser

microphone was not widely used, it is interesting to note that it is once again being manufactured for radio broadcast purposes.

Moving-coil, or Electrodynamic, Microphone.—This is a generator-type microphone, and it has proved very satisfactory for broadcasting. A diagram of a moving-coil, or electrodynamic, microphone is shown in Fig. 19. It often is called a **dynamic microphone.**

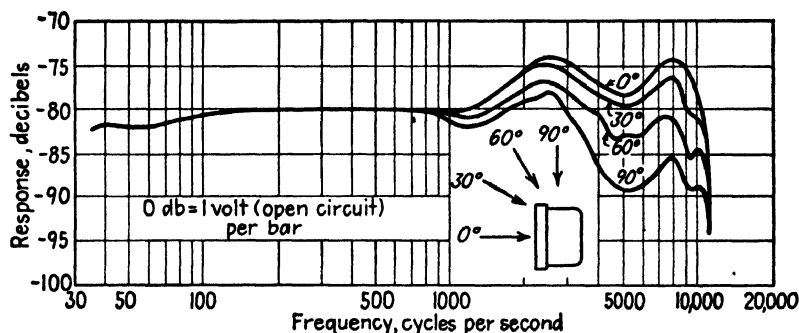



FIG. 20.—Frequency characteristics of a dynamic microphone in a *nonspherical* housing. For the meaning of the term "bar," see page 27. The microphone of Fig. 19 shows less variation than does this microphone in output with direction of arrival of the sound waves, that is, the microphone of Fig. 19 is more nondirectional.

In this device a permanent magnet produces a very strong magnetic field across a narrow air gap. A coil of wire that is attached to a diaphragm is located in this air gap. The diaphragm is caused to move back and forth in this air gap by the action of the sound waves striking it. When the diaphragm and attached coil move, the magnetic flux in the air gap, which links the coil, is changed. This induces a voltage in the coil by ordinary electric generator action, and this voltage corresponds to the sound waves that strike the diaphragm.

Various air slots and air chambers are used to influence the motion of the diaphragm and thus to control the output of the microphone. The frequency response of this microphone is excellent, and it is both quiet in operation and rugged. It is essentially nondirectional; that is, it responds quite well to sounds from any direction. Figure 20 shows the characteristics of a typical dynamic microphone in a nonspherical housing. These microphones can be located a considerable distance from the associated amplifiers if shielded cable is used.

Ribbon, or Velocity, Microphones.—These also are generator types, and are electrically similar in this respect to any other microphone that uses a magnetic field and a moving conductor to produce a voltage corresponding to the sound waves. The ribbon microphone consists of a corrugated ribbon of conducting material, such as aluminum alloy, which is hung in a strong magnetic field so that the ribbon can move under the influence of sound waves. As the sound waves move the ribbon back and forth, a voltage is induced in it, and this voltage can be amplified.

The construction of the ribbon or velocity microphone. The "corrugated" ribbon is shown between the heavy metal magnetic pole pieces. The transformer for voltage step-up and impedance-matching purposes (page 109) is in the housing at the bottom. (*Radio Corporation of America.*)

In this respect, the electrical action of the ribbon microphone is similar to that of the moving-coil type. However, from an acoustical standpoint the operation is greatly different, for the following reason: The microphones previously described were *pressure operated.* They had diaphragms, and the diaphragm was moved because the sound waves raised the air pressure on the front side of the diaphragm above the air pressure on the *enclosed* back side of the diaphragm. The ribbon microphone has no diaphragm, and both the front and back sides of the microphone are exposed to the action of the sound waves. The ribbon of this microphone is caused to move, not because of the pressure of the sound-wave front, but because the air passing through the small slits between the ribbon and the pole pieces of the magnetic circuit causes a difference in phase and a difference in pressure on the two sides of the ribbon. This action, and the voltage induced in the ribbon, is determined by the motion, or velocity, of air particles rather than by air pressure.

The frequency response of a ribbon microphone is excellent, and it is widely used, particularly for music. For close talking the basic principle of operation is such that the low frequencies are accentuated.[1] The output of a typical ribbon microphone is about −81 decibel per bar, when zero level is 0.0125 watt. In

[1] Olson, H. F., and F. Massa, "Applied Acoustics," The Blakiston Company.

most instances a step-up transformer is built into the microphone. For the microphone just considered, the internal impedance looking back into the secondary of the microphone transformer is about 250 ohms; these data are for about 1000 cycles.

The directional characteristics of a ribbon microphone are shown in Fig. 21. As indicated, this microphone is most sensitive

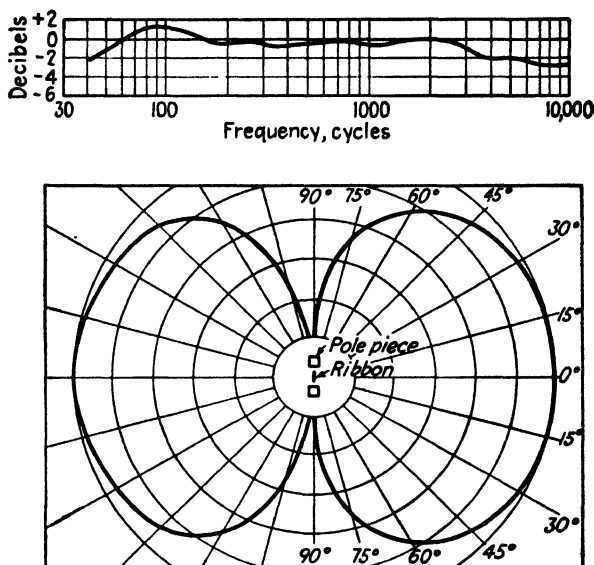



FIG. 21.—Frequency response and sound pick-up pattern for a ribbon or velocity microphone. This microphone does not respond to sound waves arriving from a source parallel to the ribbon. The lower figure shows the relative response to sounds from various directions. (*Radio Corporation of America.*)

to sounds coming from directions at right angles to the plane of the ribbon. Sounds coming from either side travel in a direction parallel to the ribbon, and they do not produce the difference of sound pressure that is necessary in order to cause ribbon motion. By suitably enclosing one side of a ribbon microphone, it can be made responsive to sounds from only one direction.[1]

Combined Moving-coil and Ribbon Microphone.—As previously explained, the pressure-operated moving-coil microphone responds almost equally well to sounds from all directions. A sound-pressure wave coming from the front moves the diaphragm in and out. A sound-pressure wave coming from the rear flows around the housing that encloses the microphone and then actuates the diaphragm. Thus the moving-coil, or dynamic, microphone is

[1] Olson, H. F., and F. Massa, "Applied Acoustics," The Blakiston Company.

nondirectional. Note in particular that for either direction of approach, a sound-pressure wave *moves the diaphragm inward.*

For the ribbon microphone, the action is quite different. If a sound wave comes from one direction, it moves the ribbon (initially of course, the sound wave being alternating in nature) in one direction, but if its direction of approach is reversed, the direction of ribbon motion also is reversed. This action is shown in Fig. 22.

Microphones that combine the two principles of operation to produce a directional response pattern are widely used. These

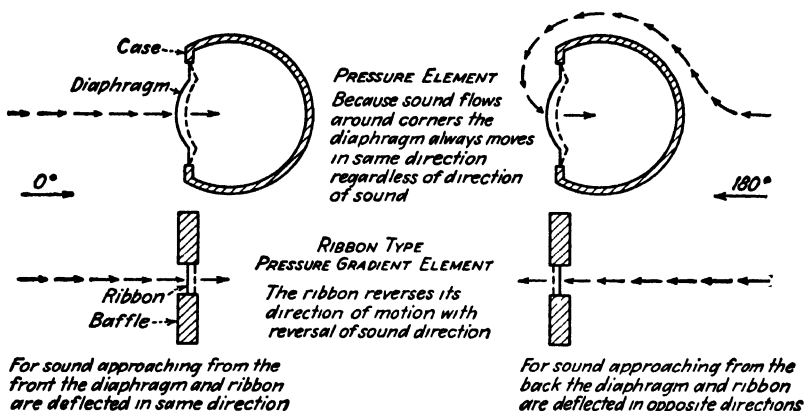



FIG. 22.—The principle of the so-called "Cardioid" microphone. This microphone combines a moving-coil element and a ribbon element. (*Western Electric Co.*)

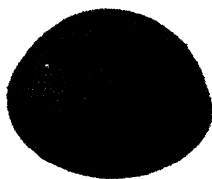
microphones contain both moving coil and ribbon elements. Thus if the moving coil and the ribbon are properly connected in series, their voltage outputs will *add* for a sound coming from the front of the microphone, but will *subtract* for sounds coming from the back of the microphone. The result is that a heart-shaped directional response pattern is obtained, giving to a popular microphone of this combined type the name Cardioid Microphone. Another advantage of this type is that an adjustment is provided so that the device can be used as a nondirectional moving-coil microphone, or as a directional ribbon microphone, or in other combinations. Directional microphones can be obtained also by acoustical means.

Crystal Microphones.—Certain crystalline substances, such as quartz and Rochelle salts, generate a voltage when the substance is deformed mechanically, as by striking or bending. This is known as **piezoelectric effect,** and this principle is used in microphones. Rochelle salts are used instead of quartz because quartz is relatively insensitive.

Suitable metallic electrodes are placed on opposite faces of a crystal, and the element is exposed to sound waves. These sound waves, although they are very feeble, are able to deform the crystal to such an extent that a voltage, corresponding to the sound waves, is created between the metallic electrodes. In one microphone, twelve of these identical elements, or "sound cells," were connected in parallel, although sometimes series-parallel combinations are used. If the elements are in parallel, a low internal impedance is obtained. If they are connected in series, or series-parallel, the internal impedance is increased and a higher output voltage is obtained.

These microphones are small, light, and have an excellent frequency response. The internal impedance of the microphone just described is largely capacitive and is about 0.02 microfarad. The output level is −67 decibels for a zero reference level, of 1 volt per bar. Most devices that use Rochelle salt crystals must be kept at temperatures below about 120°F., or they may be permanently damaged.

Telephone Receivers and Loudspeakers.—The **telephone receiver** is an electroacoustic device which is driven by electric waves and which produces sound waves that are substantially equivalent. A **loudspeaker** is a telephone receiver designed to radiate acoustic power into a room or into open air.[1] The term "telephone receiver" is used to prevent its be-

[1] American Standard Definitions of Electrical Terms, American Institute of Electrical Engineers, 1941.

A crystal microphone and its internal construction. The crystal element or sound cell is shown at the bottom. (*Brush Development Co.*)

ing confused with the word "receiver" when it is used to designate a radio-receiving set.

Telephone receivers and loudspeakers are electric motors that produce reciprocating motion instead of rotary motion. These motors are connected to some object, such as a diaphragm, that radiates energy into the air. Thus a telephone receiver or a loudspeaker consists of two parts, the **motor element** and the **acoustic radiator.**

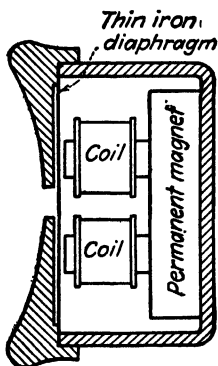



FIG. 23.—Diagram of a magnetic head receiver. The permanent magnet is circular in shape and has two centrally located soft-iron pole pieces contacting the north and south poles of the magnet. The two coils through which the signal currents pass are placed on the pole pieces.

Telephone Receivers.—These are used often in systems of radio communication. Sometimes crystal motor elements are used. In these the received and amplified electric-signal voltages are impressed on the metallic electrodes of a Rochelle salt crystal element, and these voltage variations deform the crystal, which causes sound waves to be radiated. This action is, of course, the reverse of that of the crystal microphone. These crystal receivers are light, and they have an excellent frequency response.

Most telephone receivers operate on magnetic principles, a typical receiver of this type is shown in Fig. 23. The permanent magnet attracts the thin iron diaphragm and bows it in slightly. The speech or telegraph signal currents pass through the two coils that are connected in series. For one half cycle of the signal current the magnetic effects of the coils add to that of the magnet and bow the diaphragm in further. For the opposite half cycle of the current the magnetic effects of the coils oppose that of the magnet and the diaphragm moves out slightly. In this way one cycle of signal current produces one cycle of sound.

The telephone receiver just described is a very sensitive and a very rugged device. The frequency response is not good, and distortion is produced; but this is of little consequence where information only is to be received. Such receivers are of many types and have different electrical characteristics. For instance, one type used in radio has a direct-current resistance of about 60 ohms and an inductive reactance of about 250 ohms at a frequency

of 800 cycles. Another type that is widely used has a direct-current resistance of about 1000 ohms and an inductive reactance of several thousand ohms. The impedance of a receiver varies greatly with test conditions. A fraction of a milliampere is sufficient to drive a telephone receiver. The efficiency of a telephone receiver is of the order of 1.0 per cent in converting from electric to sound waves.

Loudspeakers.—There are many types of loudspeakers, but in radio-receiving sets the moving-coil, or electrodynamic, loudspeaker is the type that is used to a large extent. A cross section of a typical moving-coil, or **dynamic, loudspeaker** (as it often is called) is shown in Fig. 24. A very strong magnetic field is produced across an air gap by a direct-current magnetizing coil or a permanent magnet.

FIG. 24.—Cross section of a moving-coil, or dynamic, loudspeaker. The magnetic field often is furnished by a permanent magnet.

Attached to a paper-cone sound radiator is a coil of a few turns of wire. This coil is situated in the air gap, and the speech or program signal currents that are to be reproduced as sounds are passed through the windings of this so-called "voice coil." By electric-motor action the alternating signal currents cause the voice coil to move back and forth in the air gap. This drives the cone, which acts as an acoustic radiator which reproduces the sound.

The direct-current resistances of the magnetizing field coils of moving-coil loudspeakers depend on the service for which the loudspeaker is designed. For instance, some field coils are designed to be energized by a 6-volt storage battery, and have a resistance of about 3 ohms. Other loudspeakers are designed so that their field coils may be used as the inductance for a rectifier filter (page 235). Loudspeaker field coils of this type have a direct-current resistance of perhaps 1000 ohms and an inductance of many henrys.

The input impedance of a loudspeaker voice coil depends on conditions of measurement, but for a typical loudspeaker it is about 8 ohms resistance and 0.0006 henry inductance at a frequency of 1000 cycles per second. The direct-current resistance of this

coil is 5 ohms. Part of this difference in resistance is due to the hysteresis and eddy-current losses that occur when alternating current is used. There is another factor involved, however. When a loudspeaker voice coil is carrying direct current, it is delivering no acoustic energy to the air. But when it is carrying alternating current, it is delivering acoustic energy to the air, and the resistance component of the input impedance must change so that it can draw this power from the source. This is similar to the ordinary electric motor in which the current changes, and thus the input impedance changes when the mechanical load connected to the shaft varies.

Sometimes an impedance-matching transformer (page 109) is used with a moving-coil loudspeaker. For the loudspeaker previously considered, the input impedance to the voice coil at 1000 cycles was 2065 ohms resistance and 0.195 henry inductance when measured through an inexpensive transformer connected between the bridge that was used in making the measurements and the voice coil.

Baffles.—If the loudspeaker of Fig. 24 is not in a cabinet, and if the cone is free to radiate in the open air, then the amount of sound that is radiated will be relatively small. The following is the reason for this: When the cone moves forward, there is a tendency to compress the air at the front and to rarefy it at the back, but this action is partly neutralized because the air particles merely flow around the edge of the cone, and cancel the pressure differences. This is quite effective in preventing the formation of sound waves. If on the other hand the loudspeaker is in a cabinet, as in a radio-receiving set, then the air particles must flow a considerable distance before neutralization can occur. Thus intense speech sounds are radiated into the room.

A similar effect is obtained if a loudspeaker is placed at or near the center of a large, flat surface called a **baffle.** This baffle should have dimensions that the shortest air path between the front and the back of the loudspeaker cone would be at least one-fourth the wavelength of the lowest note to be reproduced. With these dimensions, the sound waves produced by the front and back of the cone will not cancel.

A baffle should not be made of light material, such as thin plywood, because the ideal baffle should not vibrate. Heavy ply-

wood, or an insulating board that does not vibrate readily, should be used. If the loudspeaker is set at the center of the baffle, then the paths from the front of the cone to the back are all of the same length. Thus cancellations and reinforcements of the sounds radiated at certain frequencies may occur. If the loudspeaker is located off center, these effects are minimized. They are not particularly bothersome, however, and a loudspeaker located off center looks queer. The efficiency of a moving-coil loudspeaker in a well-designed baffle in converting electric energy to sound energy is from about 5 to 10 per cent. Sometimes horns are used on loudspeaker driving units. They serve to load the loudspeaker diaphragm to the air and also to direct the radiated sound. A good driving unit and horn have an efficiency of 30 per cent or more.

SUMMARY

Sound waves contain energy; and, once they are produced, they remain audible until the energy is absorbed (converted into heat) and the sound intensity falls below the threshold of hearing. If a sound wave is created in a room, such as a radio studio, that has insufficient sound-absorbing material present, the sound energy will be reflected back and forth for perhaps several seconds before it becomes inaudible. Such prolonged sounds will interfere with the understanding of succeeding sounds.

The reverberation time of an auditorium can be calculated by Eq. (1), $T = 0.05V/a$. For a room, such as a radio studio, where the reverberation time is in the vicinity of 1 second, this equation should be modified by subtracting from it the value $0.027V/S$. The recommended reverberation time depends on the volume of the enclosure and on the purpose for which the room is to be used.

A noise is any unwanted sound. Rooms can be kept quiet by keeping noises from being created, and by having sufficient sound-absorbing material present so that the noises which are created will be absorbed quickly. The reduction in noise level in decibels can be computed by Eq. (2), decibels = $10 \log_{10} a_2/a_1$.

The power contained in average conversational speech is about 10 microwatts. In speaking and singing, the power ratio of the strongest to the weakest sounds is about 10,000 to 1, or 40 decibels. Most sounds of speech and music are contained in a band that extends from 60 to 10,000 cycles. Orchestral music is characterized by a wider frequency range and by greater amplitude changes. The hearing range of the ear is from about 20 to 20,000 cycles for an adult.

Articulation tests have shown that information in the form of speech can be transmitted satisfactorily by a band of from about 200 to 4000 cycles. For high-quality music a band of about 50 to 8000 cycles is often used.

Listener tests show a preference for narrow-range or medium-range transmission as compared with wide-range programs.

Microphones are used to convert sound waves into electric signals. Microphones are of two general types, modifier types, such as the carbon-granule microphone, and generator types, such as the moving coil, or dynamic, microphone, the ribbon, or velocity, microphone, and the crystal microphone.

Telephone receivers and loudspeakers are in a sense electric reciprocating motors. They are composed of two principal parts, the driving, or motor, element and the acoustic radiator. The moving-coil, or dynamic, loudspeaker is the type most widely used. It should be placed in a cabinet or in a baffle for good results.

REVIEW QUESTIONS

1. Why should a person working in radio have a knowledge of acoustics?

2. What is meant by the frequency of sound waves? Are sound waves usually of a single frequency?

3. If a sound wave strikes a painted wood wall, about what per cent of the sound energy is absorbed?

4. How is sound energy absorbed by a surface?

5. How can sound transmission from one room to another be minimized?

6. Approximately what should be the reverberation time of the average radio studio?

7. Repeat Question 6 for a small auditorium.

8. Why is some small amount of reverberation advisable with music?

9. What is the number of units of sound absorption of an adult?

10. Why is a frequency-weighting network used in a sound-level meter?

11. If an office is too noisy, how can it be quieted?

12. How are speech sounds produced?

13. In Fig. 11, what are the lowest and highest frequencies of sounds produced by the instruments studied?

14. What types of instruments produce sounds of the highest frequency?

15. Name the important parts of the ear, and explain the function of each.

16. What is meant by loudness? Discuss.

17. What width of transmitted frequency band do radio listeners seem to prefer?

18. Why must the output of the generator-type microphone be very low?

19. Why are the frequency characteristics of a double-button microphone better than those of the single-button type?

20. What are the intensities of the sound waves of speech?

21. Explain the principle of operation of a moving-coil microphone. Why is it essentially nondirectional in response?

22. Explain the principle of operation of a ribbon microphone. Why is it directional?

23. Explain the principle of operation of a telephone receiver.

24. Explain the principle of operation of the moving-coil, or dynamic, loudspeaker.

25. On page 35 it states that the direct-current resistance of the voice coil of a moving-coil loudspeaker is 5 ohms and that the alternating-current resistance is 8 ohms. What accounts for this difference?

PROBLEMS

1. Repeat the calculations starting on page 9 for a similar concert room under the same conditions, except that the room is 30 feet high and 45 feet wide.

2. An office is 20 by 30 by 12 feet, the floor is covered with linoleum, and the walls are plastered. The noise level is 70 decibels, and it is desired to reduce it to 65 decibels by adding sound-absorbing materials. Calculate the amount required.

3. What would be the reverberation time of the office of Prob. 2 both before and after it is treated?

4. On page 34 it is stated that the resistance of a telephone receiver is 60 ohms and that the inductive reactance is 250 ohms at 800 cycles. If an alternating voltage of 0.5 volt of this frequency is impressed on this receiver, calculate the current and power it will draw. If the efficiency is 1.0 per cent in converting from electric power to acoustic power, calculate the acoustic-power output. Will the sound produced be audible?

5. For the loudspeaker discussed on page 35, calculate the phase angle of the voice coil, and the phase angle when measured back through the transformer secondary. (Differences in these are due to measurement errors and to the fact that a transformer is not a perfect device.) How much voltage would have to be impressed on the voice coil to have the power input be 5.0 watts? How much voltage would have to be impressed on the transformer primary so that the power input would be 5.0 watts? About how much acoustic power would this loudspeaker radiate?

CHAPTER II

ELECTRICAL FUNDAMENTALS

Speech and music were considered in the preceding chapter, so that the nature of the signals actuating a radio system would be understood. Related subjects, such as the acoustics of studios, the nature of hearing, and noise, were discussed also. In addition, microphones, telephone receivers, and loudspeakers were treated.

Radio is a specialized branch of the general field of applied electricity. Of course the same basic electrical principles apply in each of the several divisions. But in each branch there are certain viewpoints and methods that differ.

This book is designed for those who already have a knowledge of basic electricity. Nevertheless, it is advisable to review the basic electrical principles from the radio viewpoint and to introduce the radio method of attack.

Nature of Audio-frequency Signals.—Radio systems are used to transmit information that originates in the form of speech signals in radio telephony and of code signals in radio telegraphy. Radio systems are used also to transmit speech and music signals of broadcast programs and facsimile and television signals.

The electric circuits in the radio system must be able to handle the signals just listed. For this reason, the important characteristics of these signals will be summarized.

Speech.—As shown in the preceding chapter, the power ratio of the strongest to the weakest sounds is about 10,000 to 1, covering a range of 40 decibels (page 95). In speaking and singing, the voice may cover a band of from about 60 to 10,000 cycles. For speech transmission of information, tests have shown that a band of frequencies extending from 200 to 4000 cycles is adequate.

Code.—The nature of these signals depends on the type and speed of the sending apparatus; slow-speed hand operation and high-speed teletypewriter operation will give signals of different frequency characteristics. For reasons that will be apparent later, telegraph signals will be considered more in detail at the close of the next chapter. For the present, suffice it to say that a speech

channel passing a frequency band from 200 to 4000 cycles will certainly be wide enough to pass several telegraph messages.

Music.—As explained in the preceding chapter, the power ratio of the strongest to the weakest sounds in music is about 100,000 to 1, or 50 decibels (page 95). Musical sounds lie within a frequency band of from about 40 to 15,000 cycles. Tests have shown that a band of from 50 to 8000 cycles and a power, or volume, range of 10,000 to 1, or 40 decibels, is sufficient for most purposes; and the tests have also demonstrated that radio listeners apparently prefer a narrow or medium frequency range.

Facsimile and Television.—The frequency and power, or volume, ranges required, and also other important requirements of the circuits for these purposes, depend on the type of facsimile or television system. For certain facsimile systems the frequency band needed is about the same as for transmitting speech. For television, the band required is *very* wide, perhaps from 30 to 3,000,000 or 4,500,000 cycles, depending on the image perfection desired.

Excluding television, which is decidedly a special case, the signals that *actuate* a radio system exist at comparatively low frequencies and are included within the same general band of, say, 50 to 10,000 cycles. These are called **audio-frequency signals,** or merely **audio frequencies,** often abbreviated **a-f.** When the term "audio frequencies" is used, a band of from about 50 to 10,000 cycles is implied, although it may be somewhat wider or narrower. For tests or other purposes, a single frequency, such as 400 cycles or 1000 cycles, often is used. These are classed as audio frequencies because they lie within the audio-frequency band.

Nature of Radio-frequency Signals.—The distinguishing feature of radio is that the speech, music, or other signals are not transmitted through space at the frequencies at which they are produced originally. Instead, the original signals are translated or moved up to bands of higher frequencies. The signals, now existing at high frequencies, are then impressed on the sending antenna and transmitted through space. These high-frequency signals are called **radio-frequency signals,** often abbreviated **r-f.**

Radio-frequency signals are picked up by the distant receiving antenna, and enter the radio-receiving set. In this, the signals are amplified and then again translated or moved, but in this instance they are moved back down to their original frequencies, where

they are used to actuate headphones or a loudspeaker. The over-all process is shown in Fig. 25.

The question immediately arises: Why is it necessary in a radio system to raise the audio-frequency signals to higher radio frequencies for transmission through space to the distant receiving station? There are at least two reasons for this: (*a*) If it were not done, and if all radio stations transmitted the audio frequencies into space, then all radio stations would operate on the same fre-

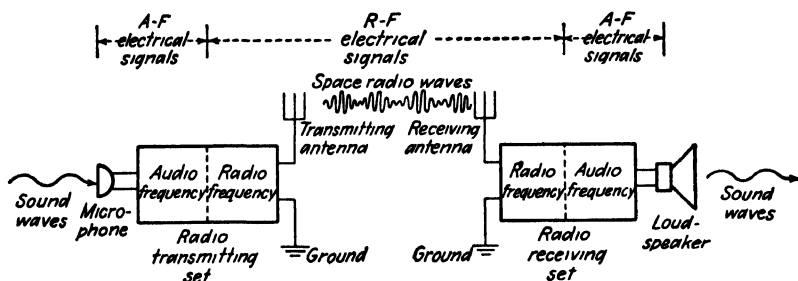



Fig 25 —A diagram of a system of radio Sound waves are converted to audio-frequency electric signals by the microphone These audio signals are raised to radio-frequency bands by the radio-transmitting set and are radiated into space by the transmitting antenna After passing through space, the radio-frequency signals are picked up by the receiving antenna and are impressed on the radio-receiving set. This set amplifies the weak radio signals and reduces them in frequency to their original audio frequencies These audio signals operate the loudspeaker, reproducing the original sound waves at the remote receiving location.

quency band, and the selection of one message and the rejection of others would be a problem; (*b*) it is difficult to radiate low-frequency signals into space, but high frequencies radiate well.

A second question is this: If audio frequencies are raised or translated to higher radio-frequency bands, what are the frequencies of these radio bands? This question must be answered in a general way. Some radio systems operate at radio frequencies as low as a few tens of thousands of cycles per second. Others operate in the regular broadcast band from about 550,000 to 1,700,000 cycles. Other stations operate at higher frequencies. The so-called "point-to-point" microwave stations operate at billions of cycles.

Thus in radio and associated applications, such as television, frequencies from zero (direct current) up to billions of cycles are used. This extreme frequency range taxes to the utmost the ingenuity of those engaged in radio. Although the same basic electrical theories apply as at any frequency, certain phenomena,

not of great consequence at low frequencies, are of much importance at the high radio frequencies.

Resistance, Inductance, and Capacitance.—These are the three basic properties of electric circuits. There are various ways of defining these properties. They will be defined and considered here entirely from the alternating-current viewpoint, the aspect of most interest in radio. Because this chapter is a review of fundamentals, the principles will be stated and in general the proofs will be omitted.

Resistance is that property of a circuit that causes a heat loss when current flows through the circuit. Sometimes the designation **effective resistance** is used. The unit of measure is the **ohm.** The power dissipated is given by the equation

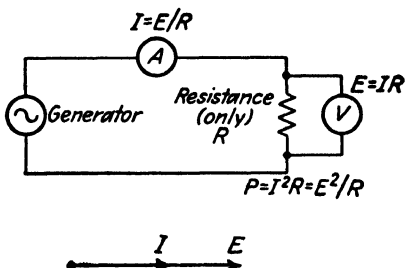



FIG. 26.—Resistance is the property of a circuit that causes heat loss. The current through a resistor and the voltage across a resistor are in phase.

$$P = I^2R, \tag{3}$$

where P is the power in **watts** that is dissipated when a current of I amperes flows through a circuit which has a resistance of R ohms. Since from Ohm's law $I = E/R$, Eq. (3) can be written

$$P = \frac{E^2}{R}, \tag{4}$$

where E is the voltage drop in volts *across the resistance*, and the other units are as previously specified. These relations are illustrated in Fig. 26.

Illustrative Problem.—The current through a 1000-ohm resistor is 15 milliamperes (0.015 ampere), and the voltage drop across the resistor as measured with a voltmeter is 15 volts. Calculate the power dissipated in the resistor.

Solution.—$P = I^2R = (0.015)^2 \times 1000 = 0.000225 \times 1000 = 0.225$ watt.

$$P = \frac{E^2}{R} = \frac{15^2}{1000} = \frac{225}{1000} = 0.225 \text{ watt.}$$

When a circuit contains resistance only, the current through the circuit is exactly in phase with the voltage drop ($E = IR$) across the circuit.

Inductance is that property of a circuit that causes the alternating current through the circuit to lag the alternating voltage across the circuit. The unit of measure of inductance is the **henry.** When an alternating current flows through an inductance coil, the alternating magnetic flux produced by this current induces a back voltage in the coil. For a circuit containing inductance only, the

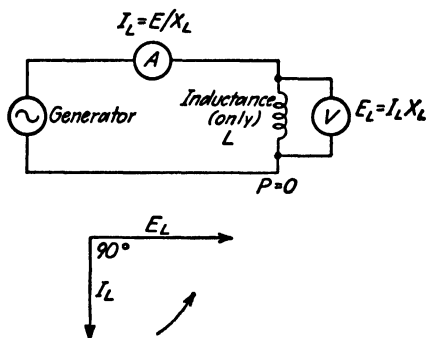



FIG. 27.—Inductance is the property of a circuit that causes the current to lag the voltage. In a circuit containing inductance only the current lags the voltage by 90°. No power is lost in a perfect inductor, but such inductors cannot be built.

magnitude of the back voltage equals that of the impressed voltage drop across the circuit. This induced back voltage is the factor limiting the current flow in the circuit. The higher the frequency, the greater will be the rate of change of current, the greater the back voltage, and the smaller the current. These relations are included in the equations

$$E_L = 2\pi f L I_L = X_L I_L \qquad \text{and} \qquad I_L = \frac{E_L}{2\pi f L} = \frac{E_L}{X_L}. \qquad (5)$$

In these expressions E is the magnitude of the back voltage in volts, which must equal the impressed voltage drop across a circuit containing inductance only of L henrys, I is the current through the circuit in amperes, f is the frequency in cycles per second, X_L is the **inductive reactance** measured in **ohms,** and $X_L = 2\pi f L$. These relations are illustrated in Fig. 27.

Illustrative Problem.—A voltage of 5.34 millivolts at a frequency of 1.2 million cycles is impressed on a coil that has an inductance of 0.01 millihenry. Neglect the resistance of the coil, and calculate the current through the coil.

Solution.—Step 1. Calculate the inductive reactance, remembering that milli means one one-thousandth.

$X_L = 2\pi fL = 6.28 \times 1.2 \times 10^6 \times 1 \times 10^{-5} = 6.28 \times 1.2 \times 10 = 75.5$ ohms.

Step 2. Calculate the current.

$$I_L = \frac{E}{X_L} = \frac{5.34 \times 10^{-3}}{75.5} = 7.07 \times 10^{-5} \text{ ampere} = 0.0707 \text{ milliampere.}$$

A circuit containing inductance *only* (such as a coil with negligible losses) cannot dissipate energy, but energy is stored in the magnetic field produced by the current flowing in an inductive circuit. When an alternating current flows in an inductive circuit, energy is stored when the current builds up, and it is returned to the circuit when the current dies out. If a circuit contains inductance only, the current I_L through the circuit will *lag* the voltage drop ($E_L = I_L X_L$) across the circuit by 90°.

Capacitance is that property of a circuit that causes the alternating current through the circuit to lead the alternating voltage across the circuit. The unit of measure of capacitance is the **farad.** When an alternating voltage is impressed across the condenser of Fig. 28, the plates of the condenser are charged alternately positively and negatively. By this action electrons flow back and forth in the line wires, and this constitutes an alternating current. If a source of alternating voltage is connected to a perfect condenser of capacitance C, the magnitude of this alternating current is

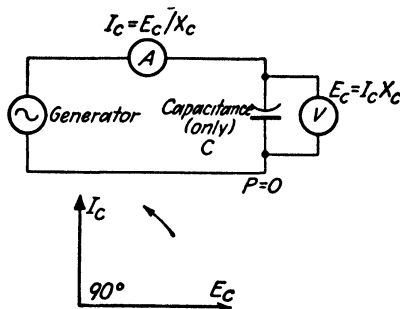



FIG. 28.—Capacitance is the property of a circuit that causes the current to lead the voltage. In a circuit containing capacitance only, the current leads the voltage by 90°. No power is lost in a perfect capacitor, and some capacitors closely approach this ideal.

$$I_C = \frac{E_C}{X_C} = E_C 2\pi fC \qquad \text{and} \qquad E_C = I_C X_C = \frac{I_C}{2\pi fC}. \quad (6)$$

In these equations I_C is in amperes, E_C in volts, f in cycles per second, C in farads, X_C is the **capacitive reactance** measured in **ohms,** and $X_C = 1/(2\pi fC)$.

Illustrative Example.—A voltage of 5.34 millivolts at a frequency of 1.2 million cycles is impressed on a condenser having a capacitance of 2,000 micromicrofarads. Assume the condenser to be perfect, and calculate the current that will flow.

Solution.—Step 1. Calculate the capacitive reactance, remembering that micro means one one-millionth.

$$X_C = \frac{1}{2\pi fC} = \frac{1}{6.28 \times 1.2 \times 10^6 \times 2000 \times 10^{-12}} = 66.3 \text{ ohms.}$$

Step 2. Calculate the current.

$$I_C = \frac{E_C}{X_C} = \frac{5.34 \times 10^{-3}}{66.3} = 8.07 \times 10^{-5} \text{ ampere} = 0.0807 \text{ milliampere.}$$

A circuit containing capacitance only (such as a perfect condenser with negligible losses) cannot dissipate energy, but energy is stored in the electric field established in the capacitance of the circuit by the impressed voltage. When the impressed voltage is alternating, energy is stored when the voltage builds up, and energy is returned to the circuit when the voltage dies out. If the circuit contains capacitance only, the current I_C through the circuit will *lead* the voltage drop ($E_C = I_C X_C$) across the circuit by 90°.

Impedance, Power, Power Factor, and Q.—The preceding section discussed resistance, inductance, and capacitance when they existed separately in circuits. This is seldom the case; they usually exist in combination. The combinations are **series circuits, parallel circuits,** and **series-parallel circuits.** A complete discussion of these combinations is not proposed at this time; certain features only will be considered (see page 72).

When an alternating voltage of E volts is connected across *any* circuit having two input terminals, a current flows into the circuit and the current will either lag, lead, or be in phase with the voltage drop across the terminals. The magnitude of the current will be

$$I = \frac{E}{Z} = \frac{E}{\sqrt{R^2 + X_e^2}}, \tag{7}$$

where Z is the **impedance** of the circuit and is measured in **ohms.** As is indicated, $Z = \sqrt{R^2 + X_e^2}$, where R is the **effective resistance** of the circuit, and X_e is the **equivalent reactance** of the circuit.

The effective resistance of a circuit was defined on page 43 and is from Eq. (3)

$$R = \frac{P}{I^2} \quad \text{or} \quad R = \frac{E^2}{P}, \tag{8}$$

where P is the total power in watts dissipated in a circuit (including all losses, such as hysteresis, eddy current, dielectric losses, radiation losses, etc.), and I is the current in amperes flowing into the circuit. The equivalent reactance X_e of a circuit is the net or

resultant reactive effects caused by the inductance and capacitance in the circuit. For a *series circuit,*

$$X_e = X_L - X_C, \tag{9}$$

as it is a convention to consider inductive reactance X_L as positive and capacitive reactance X_C as negative. This is because inductive and capacitive reactances produce opposite effects in circuits, one causing a lagging and the other a leading angle between current and voltage.

Illustrative Problem.—A resistance of 8.6 ohms, an inductance of 0.01 millihenry, and a capacitance of 2000 micromicrofarads are connected in series and across 5.34 millivolts at 1.2 million cycles. Calculate the current that will flow. The combination is shown in Fig. 29.

FIG. 29.—The series circuit of (*b*) is equivalent to (*a*) because the capacitive reactance of −66.3 ohms neutralizes all but +9.2 ohms of the reactance of the coil.

Solution.—Step 1. Calculate the reactance of the coil and condenser. This was done in the preceding section, and $X_L = +75.5$ ohms, and $X_C = -66.3$ ohms.

Step 2. Calculate the equivalent reactance of $+X_L$ and $-X_C$ in series. From Eq. (9), this is $X_e = X_L - X_C = +75.5 - 66.3 = +9.2$ ohms. Thus the circuit of Fig. 29*a* is equivalent to that of Fig. 29*b*, because having +75.5 ohms reactance in series with −66.3 ohms is the same effect as having +9.2 ohms only, since inductive and capacitive reactances act oppositely.

Step 3. Calculate the impedance of the equivalent circuit of Fig. 29*b*. From Eq. (7), $Z = \sqrt{R^2 + X_e^2} = \sqrt{(8.6)^2 + (9.2)^2} = 12.6$ ohms.

Step 4. Calculate the current that will flow. From Eq. (7)

$$I = \frac{E}{Z} = \frac{5.34 \times 10^{-3}}{12.6} = 4.23 \times 10^{-4} \text{ ampere} = 0.423 \text{ milliampere.}$$

It is indicated by Eq. (7) and the discussion that immediately follows that impedance is composed of two terms, resistance R and reactance X, and these are added at right angles. Thus impedance can be represented by a triangle as shown in Fig. 30*a*. In the impedance diagram, the angle θ is determined by the rela-

tion between the magnitudes of the resistance and the reactance, θ being the angle whose tangent equals X/R, or θ = arc tan X/R.

If each side of the impedance diagram is multiplied by the current I, then the voltage diagram of Fig. 30*b* results, and θ = arc tan IX/IR = arc tan X/R as before. This voltage diagram gives the relation between the voltages of a series circuit such as Fig. 29*b*.

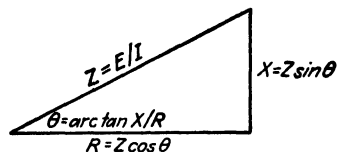
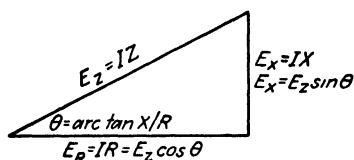
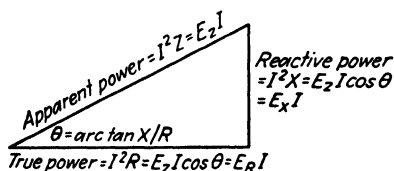



Fig. 30.—Impedance, voltage, and power diagrams for a series alternating-current circuit.

If each side of the voltage diagram is again multiplied by the current I, the power diagram of Fig. 29*c* results, the angle being θ = arc tan X/R as before. The **apparent power** represents the total power put into a circuit each cycle, the **reactive power** represents the power stored and returned, and the **true power** (or power) is the power lost in the circuit. Thus, the power dissipated in a circuit is

$$P = I^2R = EI \cos \theta, \qquad (10)$$

where E is the voltage impressed across a circuit, I is the current through the circuit, and θ is the **power-factor angle** of the circuit as determined by the magnitudes of the effective resistance R and the equivalent reactance X_e of the circuit. The term cos θ is called the **power factor,** and since the cosine of an angle varies between zero and unity, the power factor often is expressed in per cent. If the power-factor angle θ of a circuit is small, the loss in the circuit ($P = EI \cos \theta$) is large, because cos θ is large for small angles (for instance, an angle approaching zero). If the power-factor angle θ is large, the loss in the circuit is small, because cos θ is small for a large angle (that is an angle approaching 90°). The product of power and time is **energy,** usually measured in **watt hours** or **kilowatt hours.**

In the field of radio, the factor **Q** is used to indicate if the loss in a circuit is low or high. By definition,

$$Q = \frac{X}{R}, \tag{11}$$

and Q is termed the **energy-storage factor** of the circuit. Thus, if the equivalent reactance X_e of a circuit is high compared with the effective resistance R, the ratio of the energy stored per cycle to the energy dissipated per cycle will be high because $Q = X_e/R$ will be a large number. In a circuit that has a low Q, much of the apparent power entering the circuit will be dissipated in the resistance. Another term used in radio is the **dissipation factor,** designated by **D.** This is the opposite, or reciprocal, of the storage factor

$$D = \frac{1}{Q} = \frac{R}{X}. \tag{12}$$

The term Q largely is used when discussing coils and circuits; the term D often is used in discussing condensers.

Illustrative Problem.—For the circuit of Fig. 29*b*, calculate the values for the three triangles of Fig. 30.

Solution.—Step 1. From the values in the preceding illustrative example, $Z = 12.6$ ohms, $R = 8.6$ ohms, and $X_e = 9.2$ ohms. It remains to find θ. As indicated in Fig. 30*a*, $\theta = \text{arc tan } X_e/R = \text{arc tan } 9.2/8.6 = 1.07$, and $\theta = 47°$ (approximately). As a further example of the use of Fig. 30*a*, $\sin 47° = 0.731$, and $\cos 47° = 0.682$. Thus if θ and Z are known, $X_e = Z \sin \theta = 12.6 \times 0.731 = 9.2$ ohms, and $R = Z \cos \theta = 12.6 \times 0.682 = 8.6$ ohms (approximately).

Step 2. For Fig. 30*b*, the total voltage drop across the circuit will be $E_Z = IZ$. From the preceding illustrative example, $Z = 12.6$ ohms, and $I = 4.23 \times 10^{-4}$ ampere. Then $E_Z = 12.6 \times 4.23 \times 10^{-4} = 5.34 \times 10^{-3}$ volt. The voltage drop across the coil and condenser (it is these together that give the equivalent reactance) is $E_X = IX_e = 4.23 \times 10^{-4} \times 9.2 = 3.89 \times 10^{-3}$ volt. The voltage drop across the resistor will be $E_R = IR = 4.23 \times 10^{-4} \times 8.6 = 3.64 \times 10^{-3}$ volt.

Step 3. For Fig. 30*c*, the apparent power is $E_Z I = 5.34 \times 10^{-3} \times 4.23 \times 10^{-4} = 2.26 \times 10^{-6}$ volt-ampere. The reactive power is $E_Z I \sin \theta = 2.26 \times 10^{-6} \times 0.731 = 1.65 \times 10^{-6}$ volt-ampere. The true power is $E_Z I \cos \theta = 2.26 \times 10^{-6} \times 0.682 = 1.54 \times 10^{-6}$ watt.

Step 4. From the calculations of Step 1, the power-factor angle $\theta = \text{arc tan } X_e/R = \text{arc tan } 9.2/8.6 = \text{arc tan } 1.07$, and $\theta = 47°$ and $\cos \theta = 0.682$. Thus the power factor is 68.2 per cent. The Q of the circuit $Q = X_e/R = 9.2/8.6 = 1.07$, and the dissipation factor $D = 1/Q = 0.936$.

Resistance of Conductors.—The equation used to calculate the resistance of a conductor is

$$R = \frac{\rho l}{a}, \tag{13}$$

where R is the resistance in ohms, when ρ is the resistivity of the material in ohms per centimeter cube, l is the length of the conductor in centimeters, and a is the cross-sectional area in square centimeters. Of course this equation may be used also to calculate the resistance of resistors as well as conductors, porvided that ρ is the resistivity of the material of the resistor.

Illustrative Problem.—Calculate the resistance at 20°C. of a copper bar 2 inches by 3 inches in cross section and 9 feet long

Solution—First convert all dimensions to centimeters. Then the dimensions are 2 × 2.54 = 5 08 centimeters by 3 × 2 54 = 7 62 centimeters (the cross-sectional area is 5 08 × 7 62 = 38 7 square centimeters) and the length is 9 × 12 × 2 54 = 274 centimeters.

$$R = \frac{\rho l}{a} = 1.724 \times 10^{-6} \times \frac{274}{38.7} = 12.2 \times 10^{-6} \text{ ohm.}$$

The value of ρ was obtained from Table II

TABLE II.—ELECTRICAL CHARACTERISTICS OF METALS AND ALLOYS USED IN RADIO[1]

Based on a temperature of 20°C.

Material	Composition	ρ microhms per centimeter cube	ρ' ohms per circular-mil-foot	α
Aluminum	In wires	2 828	17 01	0 00403
Copper	Standard annealed	1 724	10 37	0 00393
Copper	Hard-drawn No 12 A W.G	1 772	10 68	0 00382
Iron.		9 8	59 0	0 006
Nickel		10 0	60 0	0 005
Silver.		1 629	9 8	0 00381
Brass	Commercial	2 4	25	0 002
Advance	Cu-Ni	49	295	±0 00002
Manganin	Cu-Mn-Ni	48	290	0 00001
Nichrome V	Ni-Cr	93	650	0 00013

[1] These data were compiled to a great extent from such electrical and radio handbooks as "Standard Handbook for Electrical Engineers" McGraw-Hill Book Company, Inc, "Radio Engineers' Handbook" McGraw Hill Book Company Inc and "Electric Communication and Electronics" Vol 5 of "Electrical Engineers Handbook' John Wiley & Sons Inc It is pointed out that such values depend on exact chemical composition purity of specimen, mechanical and heat treatment, etc

Resistance of Wires.—The resistance of wires of circular cross-sectional area is given by the equation

$$R = \frac{\rho' l}{a}, \tag{14}$$

where R is the resistance of the wire in ohms, ρ' is the resistance of the material in ohms per circular-mil-foot, l is the length of the wire in feet, and a is the cross-sectional area of the wire in circular mils. This circular-mil area equals the wire diameter expressed in thousandths of an inch and squared.

Illustrative Problem.—Find the resistance at 20°C. of a hard-drawn copper wire that is 0.0808 inch in diameter and 1000 feet long.

Solution.—A diameter of 0.0808 inch = 80.8 mils. Then d^2 = 6530 circular mils. From Table II, ρ' = 10.68 ohms. Then

$$R = \frac{\rho' l}{a} = 10.68 \times \frac{1000}{6530} = 1.64 \text{ ohms.}$$

Wire sizes and other information for standard annealed copper (soft) are shown in Table III. This is the type of wire used for winding coils and for wiring. Hard-drawn copper wires are used for antennas and transmission lines, because these have much higher tensile strength than annealed or soft copper. The resistance of a hard-drawn wire of about No. 12 A.W.G. is about 3 per cent higher than a similar annealed or soft copper wire. All wire data given in this chapter are for direct current, and they also apply with little error at audio frequencies. Information for radio frequencies will be given in Chap. V.

Resistance Variations with Temperature.—The resistance of a wire varies with temperature in accordance with the equation

$$R_2 = R_1[1 + \alpha(t_2 - t_1)], \tag{15}$$

where R_2 is the final resistance at temperature t_2, R_1 is the resistance at temperature t_1, and α is a temperature coefficient of resistance. The value of α usually is given for a starting temperature of 20°C., and the value of R_1 is then the resistance at this temperature. The values of t_1 and t_2 are in degrees centigrade.

Illustrative Problem.—A transmission line composed of two hard-drawn No. 12 copper conductors is 250 feet long (total wire length 500 feet). What will be the resistance (to direct current and at audio frequencies) if the temperature is 30°C.?

Solution.—From the preceding illustrative problem the resistance will be 1.64/2 = 0.82 ohm at 20°C. From Table II α = 0.00382 for hard-drawn copper. At 30°C. the resistance will be

$$R_2 = 0.82\,[1 + 0.00382\,(30 - 20)] = 0.88 \text{ ohm.}$$

Insulators and Dielectrics.—Materials that have very high resistances are used for insulating one portion of a circuit from

TABLE III.—WIRE TABLE FOR INTERNATIONAL STANDARD ANNEALED COPPER
American Wire Gauge (B. & S.)

B. & S. gauge, No.	Diameter, mils, d	Area, circular mils, d^2	Ohms per 1000 ft. at 20°C., or 68°F.	Pounds per 1000 ft.
0000	460.00	211,600	0.04901	640.5
000	409.64	167,810	0.06180	508.0
00	364.80	133,080	0.07793	402.8
0	324.86	105,530	0.09827	319.5
1	289.30	83,694	0.1239	253.3
2	257.63	66,373	0.1563	200.9
3	229.42	52,634	0.1970	159.3
4	204.31	41,742	0.2485	126.4
5	181.94	33,102	0.3133	100.2
6	162.02	26,250	0.3951	79.46
7	144.28	20,816	0.4982	63.02
8	129.49	16,509	0.6282	49.98
9	114 43	13,094	0.7921	39.63
10	101.89	10,381	0.9989	31.43
11	90.742	8,234.0	1.260	24.93
12	80.808	6,529.9	1.588	19 77
13	71.961	5,178.4	2.003	15.68
14	64.084	4,106.8	2.525	12.43
15	57.068	3,256.7	3.184	9.858
16	50.820	2,582.9	4.016	7.818
17	45.257	2,048.2	5.064	6.200
18	40.303	1,624.3	6.385	4.917
19	35.890	1,288 1	8.051	3.899
20	31.961	1,021.5	10.15	3.092
21	28.462	810.10	12.80	2.452
22	25 347	642.40	16.14	1.945
23	22 571	509 45	20 36	1.542
24	20.100	404.01	25.67	1.223
25	17.900	320 40	32.37	0.9699
26	15.940	254 10	40.81	0.7692
27	14.195	201.50	51.47	0.6100
28	12.641	159 79	64.90	0.4837
29	11.257	126 72	81.83	0.3836
30	10.025	100.50	103.2	0.3042
31	8.928	79.70	130.1	0.2413
32	7.950	63.21	164.1	0.1913
33	7.080	50.13	206.9	0.1517
34	6.305	39.75	260.9	0.1203
35	5.615	31.52	329.0	0.0954
36	5.000	25.00	414.8	0.0757
37	4.453	19.82	523.1	0.0600
38	3.965	15.72	659.6	0.0476
39	3.531	12.47	831.8	0.0377
40	3.145	9.89	1049	0.0299

another and for the dielectric between the plates of a condenser. In a sense, insulators and dielectrics are the same thing; the relation is somewhat as follows: If a material is used to separate two parts of a circuit, it is being used as an insulator. But when it is used in this manner, often it has a strong electric field established in it because the two parts of a circuit are at a difference of potential. The properties of the material as a dielectric must then be considered, because under these conditions a power loss will occur in the "insulator" just as a loss occurs in the dielectric of a condenser.

The resistance of an insulator is controlled by two factors; these are the *surface* leakage and the leakage through the body, or the *volume* leakage. There are so many variable factors (such as humidity) that enter into leakage, and insulators are of so many different shapes, that often it is difficult to calculate the resistance of insulators. Also, in general, the leakage is not the most important measure of insulator performance, particularly at high radio frequencies; for instance, the dielectric loss may be the determining factor. As an indication of the laws governing insulator leakage, the surface leakage is given by the equation

$$R_S = \frac{\text{surface resistivity} \times l}{w}, \tag{16}$$

where the surface resistivity is expressed in ohms per centimeter square and is of the order of 8×10^{15} ohms for Bakelite and 2×10^{13} ohms for Steatite both at 50 per cent relative humidity, l is the length of the leakage path, and w is the width of the path, both in centimeters. The volume resistance of an insulator can be determined by using Eq. (13), where ρ (now the volume resistivity) in ohms per centimeter cube is about 2×10^{16} ohms for Bakelite and 2.5×10^{14} ohms for Steatite.

The discussion just given applies for direct current and for low-frequency alternating current, and it applies at radio frequencies if the surface and volume resistivities remain constant at the values given, or if correction factors are known for these higher frequencies. At radio frequencies, probably the most important electrical characteristic of insulators and dielectrics is the **dielectric hysteresis loss.**

As an example of dielectric hysteresis loss, consider a layer of some dielectric, such as Bakelite, between the plates of a condenser on which a high voltage of radio frequency is impressed.

This rapidly alternating voltage establishes an electric field in the dielectric. Because the voltage is rapidly changing, this field exists first in one direction and then in the other. The rapid reversing of this field causes a heat loss, called a dielectric hysteresis loss, in the insulating material used as the dielectric between the condenser plates. This loss is very great at high frequencies, increasing directly as the first power of the frequency, because a given loss occurs each cycle. Thus plastic materials (such as the ones used in tube bases) may become soft, wood may be charred, and materials containing moisture may steam. There are many practical uses made of this principle; for instance, plastics are heated for molding, rubber is vulcanized, plywood glue is set, and foods are dried, cooked, and sterilized by high-frequency electric fields.

The electrical characteristics of certain materials used for insulators and dielectrics are shown in Table IV.

TABLE IV.—ELECTRICAL CHARACTERISTICS OF MATERIALS USED FOR INSULATORS AND DIELECTRICS IN RADIO[1]

Material	Dielectric constant	Power factor, per cent — Cycles per second: 60	1000	10^6	Machine-ability
Cellulose acetate	6–8	7	....	3–6	Very good
Cellulose nitrate	4–7	5–9	5	5	Very good
Fiber	4–5	6–9	5	5	Very good
Glass (Pyrex)	4.5	...	0.5	0.2	Very poor
Mica (clear India)	7–7.3	0.03	0.02	0.02	
Bakelite (low loss)	5.3	2.5	1.4	0.7	Poor
Porcelain (wet process)	6.2–7.5	2	1	0.7	Very poor
Rubber (hard)	2–3	1	1	0.5–0.9	Fair
Steatite	6.1	1	0.4	0.3	Very poor
Styrene (polymerized)	2.4–2 9	0.02	0.02	0.03	Good

[1] From *General Radio Experimenter*, June, 1939.

All other factors being equal, the materials that have a high dielectric constant would be good for the dielectrics of condensers, because they would give a large capacitance for given electrode area. Those which have a low loss, as indicated by a low power factor, would be good at very high frequencies, where other materials might develop a high temperature.

As an illustration of the use of Table IV, Bakelite at 1,000,000

cycles has a power factor of 0.7 per cent, and from the discussion on page 48 this corresponds to a power-factor angle of θ = arc cos 0.007 = 89.6°. At this same frequency, the power-factor angle for Styrene is θ = arc cos 0.0003 = 89.99°. Thus if these two materials were used for the dielectrics in otherwise identical condensers, the loss in the condenser with Styrene dielectric would be less than the loss in the condenser with the Bakelite dielectric, because (all other factors being assumed equal) the current that flowed into the first condenser would be more nearly 90° out of phase with the voltage across the condenser. Power cannot be absorbed by a circuit if the current and voltage are 90° out of phase.

Fig. 31.—A resistor has unwanted inductance in series and unwanted capacitance in parallel. These may be troublesome at radio frequencies.

Circuit Elements.—Electric circuits often are composed of circuit elements or units that have primarily resistance, inductance, or capacitance. Such circuit elements correctly are called **resistors, inductors,** or **capacitors.** The terms "coils" and "condensers" also are proper for these last two elements. The term "resistance" to designate a resistor is falling into disfavor. These circuit elements are connected in series or in parallel combinations.

To a great extent the operation of a radio circuit depends on the frequency characteristics of the resistors, inductors, and capacitors used. It was mentioned that these contained primarily resistance, inductance, or capacitance. However, small values of unwanted properties may exist in an element; for example, an inductor may have appreciable capacitance between turns. It is of much importance in radio that the electrical characteristics of the individual elements are known.

Resistors Used in Radio Circuits.—The resistors used are of two basic types, wire-wound resistors and composition resistors. These resistors may be fixed in resistance, or may be variable. A resistor often has unwanted inductance in series and unwanted capacitance in parallel, as Fig. 31 indicates. Whenever an electric current flows, it creates a magnetic field, and this magnetic field causes the unwanted inductive effect. The unwanted capacitance exists between the metal terminals, between parts of the resistor

and the terminals, and between the parts of the resistor, just as capacitance exists between any metal parts.

Effort is made in the design of resistors to keep the inductance and capacitance effects shown in Fig. 31 as low as practicable. At audio frequencies these effects often are negligible, but at radio frequencies they become of much importance These sometimes

FIG 32—These four wire-wound resistors have the same resistance values, but their sizes give them different power-dissipating capabilities and hence different current-carrying capacities From top to bottom the power ratings are 100, 50, 20, and 10 watts. (*International Resistance Co*)

are called "residual" effects. Because wire-wound resistors and composition-type resistors are so different, they will be considered separately.

Wire-wound Resistors.—Those used in radio apparatus usually are wound on a ceramic form, and the wires often are covered with a baked-on vitreous enamel or other suitable coating that protects the wires mechanically and insulates the resistor electrically. Such resistors are available in many resistance values up to about 100,000 ohms and in power-handling capacities up to about 200 watts, although, of course, higher values are possible and are made. Resistors of this type are shown in Fig. 32.

The unwanted residual inductance of a wire-wound resistor can be minimized in several ways. In general, inductive effects are caused by the magnetic flux a current produces. Thus inductance can be minimized by so winding a resistor that little flux is produced. There are two simple ways of doing this, as will now be explained. (*a*) If the resistance wires are wound on a very thin card (often mica is used), then the area of the "core" or the "coil" will be small, and from basic magnetic theory the flux produced and hence the inductance will be small. (*b*) If the resistance wires are so wound that adjacent turns are close together and carry the same current in *opposite* directions, the magnetic flux-producing tendencies will cancel to a large extent, and but little flux and inductance will result. There are several ways of accomplishing method (*b*), as shown in Fig. 33. Wire that has little change in resistance with temperature variations is used for winding resistors (page 51).

FIG. 33.—Three of the possible methods of winding resistors so that self-inductance is minimized. In each the principle is the same; the currents in adjacent wires are in opposite directions, and the magnetic effects cancel.

The unwanted residual capacitance can be kept low by using small terminals, by separating the wires as far apart as practicable, and by arranging the wires so that there is little difference of potential between adjacent turns. This last feature is effective because if adjacent wires are at essentially the same potential then there will be little current flow through the capacitance between them and hence the over-all capacitive effect in the resistor will be low. It will be noted that the remedy for residual inductance may increase the residual capacitance.

Wire-wound resistors are used generally for direct current and in audio-frequency circuits. They seldom are used at radio frequencies, except at the lower end of the radio spectrum, or where the residual inductance and capacitance are of no consequence. Measurements made on a 10-watt 500-ohm wire-wound resistor of the type shown in Fig. 32 showed a resistance of 503 ohms and an inductance of 0.0021 henry at 1000 cycles. Measurements made on a 1000-ohm wire-wound voltage divider of the type widely used in laboratories for experimental

purposes (see Fig. 38) showed a resistance of 984 ohms and an inductance of 0.0045 henry. This voltage divider had a maximum current-carrying capacity of 110 milliamperes. These measurements also were made at 1000 cycles. A much higher frequency would be necessary to study the effect of the shunting capacitance of Fig. 31. An accurate wire-wound resistor with little residual inductance and capacitance is shown in Fig. 34.

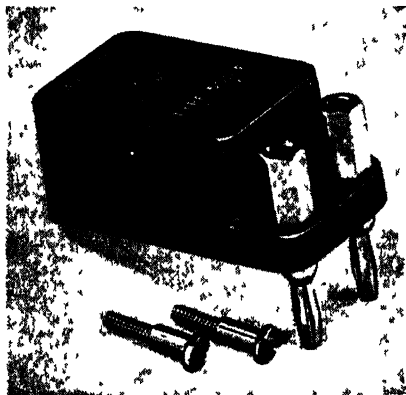



FIG. 34.—An accurate wire-wound noninductive resistor used in radio-frequency bridge measurements. (*General Radio Co.*)

Composition-type Resistors.—These often are called **carbon resistors** because they are made of current-conducting material, such as powdered carbon or graphite mixed with some nonconductor and a suitable binder in the proper proportion to give the resistance desired. Typical radio resistors are shown in Fig. 35. Such resistors are made in several power-handling capacities with various

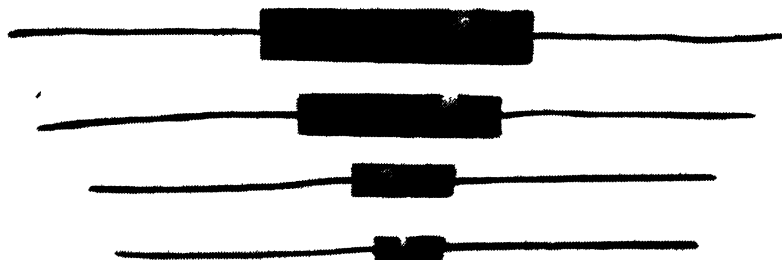



FIG. 35.—Resistors used in radio of different power-handling capacities, ranging from 2 watts to ¼ watt. These are of the metallized type, and commonly are made in resistances from millions of ohms down to a few hundred ohms.

arrangements of terminals. The bodies of these resistors usually are insulated so that they will not be in electric contact should they touch.

For convenience the so-called **metallized resistor** is classed here as a "composition type." In these a suitable conducting metallic film is painted or deposited on a glass rod or other insulator. This portion is then provided with suitable connecting leads, and the body is insulated.

Resistor Markings.—The value of power-handling capacity and the resistance of composition-type resistors often is stamped on the resistor. They are sometimes color coded in addition, but sometimes only the color code is marked on the resistor. The color code used for resistors is shown in Table. V.

TABLE V.—RESISTOR COLOR CODE[1]

Color	Significant figure	Decimal multiplier	Tolerance, per cent
Black	0	1	
Brown.....	1	10	1
Red.......	2	100	2
Orange....	3	1,000	3
Yellow	4	10,000	4
Green.....	5	100,000	
Blue	6	1,000,000	
Violet.....	7	10,000,000	
Gray......	8	100,000,000	
White....	9	1,000,000,000	
Gold......	...	0.1	5
Silver	....	0 01	10
No Color .	.. .	..	20

[1] Radio Manufacturers Association, 1938 Standard as modified.

These colors are shown as bands and dots on the resistor. For the axial-lead resistor of Fig. 36*a*, which is a type often encountered, the bands are designated *A*, *B*, *C*, and *D* as shown. For this resistor, band *A* indicates the first significant figure, band *B* the second significant figure, band *C* the decimal multiplier, and band *D* the tolerance. Thus, a 100,000-ohm resistor of 10 per cent tolerance would be marked as follows: Band *A*, brown; band *B*, black; band *C*, yellow; band *D*, silver. A 150,000-ohm resistor would be marked the same, except that band *B* would be green. A 1-megohm resistor of 10 per cent tolerance would be marked brown, black, green, and silver. For the radial-lead resistor of

Fig. 36*b*, the markings are as shown. The color code of Table V applies.

The per cent tolerance designates the percentage variation in resistance that may be found in resistors of a given group. For example, if a 100,000-ohm resistor is marked with a gold band, the resistance may be anywhere between 105,000 and 95,000 ohms; if it is marked with a silver band, the actual resistance may be as high as 110,000 ohms, or as low as 90,000 ohms.

The resistance of carbon resistors varies with temperature as much as 5 per cent over the normal operating range, depending on the type, size, and make. It is also caused to vary slightly by other factors, such as humidity and age. Carbon resistors are likely to be "noisy" (that is, cause random current variations) because the granule to granule path will vary slightly as it carries current. This is particularly true for carbon resistors of high value, and for other types of resistors as well, because in such resistors the material used may be regarded as a semiconductor and in such the path is variable. The resistance value at which such factors become of importance depends in a large measure on conditions; for example, carbon resistors of high value may cause a high-gain amplifier to be noisy, but could be used with no difficulty in a low-gain circuit. Perhaps it is reasonable to say that resistors of 0.5 megohm and above may cause trouble in this respect in some circuits. Resistors carrying relatively large direct currents in addition to alternating currents may be noisy because of the heating effect of the direct current.

Fig. 36.—Method of color coding small fixed resistors. See Table V and the discussion accompanying. Other codings have been used Sometimes, for the radial-lead type (b), marking *A* extends over the end shown as *D*, marking *C* is a band within *A*, and marking *D* is omitted.

The inductance of the composition-type resistor is very small, and is in fact negligible except at very high frequencies, say above 10,000,000 cyles for most applications. The inductance is comparable in magnitude with that of a piece of wire of similar length. Capacitance exists between leads, terminals, and the body of the insulator and between granules. The effect of the capacitance is to reduce greatly the resistance at very high frequencies and to

introduce capacitive reactance. Metallized resistors do not show such large variations in this respect.[1]

Variable Resistors.—Wire-wound resistors can be made variable by arranging suitable taps at various points, or by having a contact that slides along the resistor wire. The sliding-contact type is popular in radio apparatus.

If only two contacts are brought out as indicated in Fig. 37*a*, the device is a **rheostat.** If, on the other hand, three contacts are brought out as in Fig. 37*b*, the device is a **voltage divider.** This is often called a "potentiometer," a term that is not in accordance with the American Standard Definitions of Electrical Terms. A typical wire-wound voltage divider is shown in Fig 38.

FIG 37—A rheostat (*a*) has only two terminals, but a voltage divider (*b*) has three terminals available.

When a voltage divider is of the composition type instead of being wire wound, it often is called a **volume control,** and as such it is used to vary the input to, and, hence, the output from, an amplifier or similar device. Such volume controls are of several kinds, and consist of a conducting coating deposited on a suitable form. The "sliding" contact used must be of a special type to prevent damage to the coating. A roller or a "floating" disk are examples of such contacts. With the floating-disk method, arrangements are made so that the disk presses down at the point desired, and no contact sliding on the conducting surface is required. A typical volume control has a resistance of 250,000 ohms and a power-handling capacity of one watt.

FIG. 38.—A wire-wound voltage divider very useful at audio and the lower radio frequencies (*General Radio Co*)

[1] Extensive data on this and similar subjects will be found in the various technical journals, and summaries of such data are given in the various radio handbooks, such as those listed on p. 50.

Inductors.—These are used in radio apparatus in a variety of forms. They can be divided into two groups: those for direct-current and audio-frequency circuits and those for radio-frequency circuits. In general, those for direct current and audio frequencies have cores of magnetic material, and those for radio-frequencies have air cores. There are exceptions to this, however, because magnetic cores are used in some radio-frequency inductors.

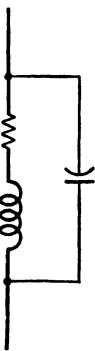



Fig. 39.—An inductor has unwanted series resistance in the wire of the coil and unwanted distributed capacitance between coil turns.

Because of the resistance of the wire of the coil, and because of capacitance between the turns of the coil, an inductor has residual resistance and capacitance, as indicated in Fig. 39. In general, in the design of inductors care is taken to keep the resistance and capacitance as low as possible. Residual capacitance acts to neutralize some of the inductance, and it will cause an inductor to be resonant (page 80) at certain frequencies. Low resistance is desired so that a coil will have a high Q, that is, a high ratio of inductive reactance to resistance. From Eq. (11), $Q = X/R = \omega L/R = 2\pi fL/R$ for an inductor.

Inductors are used for many purposes. For instance, they have the following uses: (*a*) As circuit elements, or "chokes," that offer high impedances to high frequencies but low impedences to low frequencies and direct current; (*b*) in conjunction with capacitors to form resonant circuits; and (*c*) to cause currents to lag voltages, that is, to produce phase shifts.

Inductors with Magnetic Cores.—As mentioned in the preceding section, inductors with magnetic cores are used largely in direct current and audio-frequency circuits, but sometimes are used at radio frequencies.

For direct currents and audio frequencies a typical inductor is made as shown in Fig. 40. The magnetic core usually is of silicon-steel laminations. The coil, composed of many turns of fine copper wire, often is placed on the center portion of the magnetic core. A typical inductor, such as the kind used for a "choke coil" in a small radio power supply (a device that converts from alternating to direct current), has 12 henrys inductance at 200 milliamperes, and a direct-current resistance of 200 ohms. Note that it was stated that the coil had *12 henrys inductance at 200 milli-*

amperes. This statement is stressed, because when a coil carries both direct and alternating current (as it does in a power supply), the amount of direct current fixes the point of operation on the magnetic saturation curve of the coil. For this reason the inductance, and inductive reactance, offered to the alternating current varies with the amount of direct current flowing (see page 245). The inductance of such a coil is more properly specified as an **incremental inductance.**

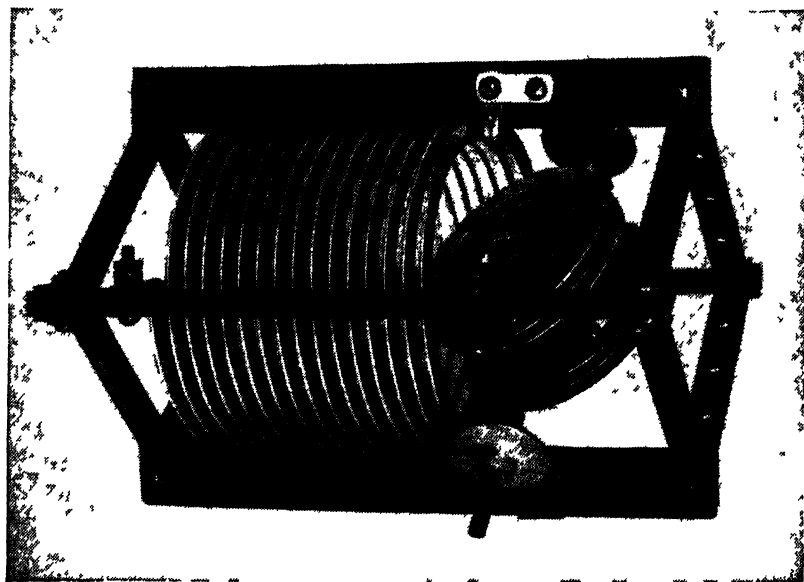
When coils with cores of magnetic material, such as silicon-steel laminations, are used in alternating-current circuits, hysteresis losses and eddy-current losses occur in the laminations. Theoretically, hysteresis loss increases directly as the first power of the frequency, and eddy-current loss increases as the frequency *squared.* This last fact limits the use of magnetic cores at radio frequencies. By making the laminations very thin, the eddy-current loss can be kept within reason at audio frequencies, and at the low radio frequencies.

FIG. 40.—An inductor or "choke coil" like that used in a power-supply filter to provide direct-current power for operating vacuum tubes.

In some audio-frequency coils and transformers, and in cores for coils and transformers used at radio frequencies, the magnetic material is powdered, mixed with a suitable binder, and then compressed at high temperature into a solid mass. This effectively insulates one metallic particle from another so that the eddy-current loss is not prohibitive even at radio frequencies of the order of 10,000,000 cycles and higher. Magnetic alloys such as the Permalloys, are used in this way. In some radio-frequency coils and transformers the inductance is varied by sliding a small part of the magnetic circuit (such as a small slug of the compressed powdered material) in and out of the core of the coil. The design of inductors with magnetic cores is complicated by the fact that the magnetic characteristics vary and that curves must be used.

The methods of making such calculations are covered in handbooks, such as the ones listed on page 50.

Inductors with Air Cores.—By far the majority of the coils used at radio frequencies are of this type, the reason being, of course, that the losses in magnetic cores are, generally speaking, quite high. At radio frequencies the capacitance between turns becomes very important; in fact, if the frequency is sufficiently high, more current may flow through the distributed capacitance between turns than is flowing through the turns of wire. If this is true, the inductor will draw a leading current, and will be, in effect, a capacitor. For this reason, a high-inductance audio-frequency inductor, such as the one shown in Fig. 40, would not be effective at radio frequencies.

FIG 41 —A typical inductor or "choke" used in radio-frequency circuits The coil is wound and is sectionalized so that the unwanted distributed capacitance is minimized (*Bell Telephone Laboratories*)

An air-cored coil or inductor, such as is used in radio-transmitting equipment (*E F. Johnson Co*)

A typical radio-frequency inductor, or "choke," is shown in Fig. 41. The windings are sectionalized, so that in effect several small inductors (with low distributed capacitance between turns) are connected in series. In this way, less residual capacitance results than if the wires are all close together. The design of radio-frequency air-cored coils is a very involved subject because of the uncertainty of the exact magnetic path and because of the great variety of coils used. Here again, reference must be made to handbooks such as the books listed on page 50. Variable air-cored inductances can be made by varying the tap on a coil. There are other ways of obtaining variable inductance, as will be explained on page 104.

Types of capacitors used in radio equipment. At the top is shown a small electrolytic capacitor. Next in order are a paper capacitor, three mica capacitors, and last, two ceramic capacitors.

Capacitors.—Many types of capacitors are used in radio equipment, and as for inductors, these capacitors can be classified as (*a*) those for direct-current circuit and audio-frequency circuits and (*b*) those for radio-frequency circuits. Those for direct current and audio frequencies include the metal-foil type (with a paraffined, waxed, or oiled paper as the dielectric) and the electrolytic type. For radio frequencies, air condensers, or condensers with mica or some other material having a low dielectric hysteresis loss are used.

It is usually quite easy to calculate the capacitance of most parallel-plate capacitors by using the equation

$$C = \frac{8.842\,KA}{10^{14}d}, \tag{17}$$

where C is the capacitance in farads, when K is the dielectric constant (Table IV), A is the area of the dielectric (see illustrative

problem) in square centimeters, and d is the thickness of the dielectric in centimeters. This equation applies when the plates are close together as in the usual radio capacitor.

Illustrative Problem.—A condenser is arranged as shown in Fig. 42. There are five plates, each having an area of 25 square centimeters, and the dielectric is of Pyrex glass 0.16 centimeters thick. Assume that the glass occupies all the area between the plates, and calculate the capacitance.

FIG. 42.—In a multiplate condenser it is the total area of the dielectric that determines the capacitance.

Solution.—The cross-sectional area of the dielectric parallel to the plates is the factor determining the capacitance. This area is 25 × 4 = 100 square centimeters. From Table IV the dielectric constant of Pyrex glass is 4.5. From Eq. (17)

$$C = \frac{8.842 \times 4.5 \times 100}{10^{14} \times 0.16} = \frac{249 \times 10^{2}}{10^{14}} = 249 \times 10^{-12} = 249 \text{ micromicrofarads.}$$

If a dielectric having a low power factor (Table IV) is used, then the current entering a capacitor will be almost 90° out of phase with the voltage across the capacitor, and little power will be dissipated in it. This is expressed in radio by saying that the Q (which is X_C/R) is high, or that the dissipation factor D (which is $1/Q$ or R/X_C) is low. The expression for the capacitive reactance is $X_C = 1/(2\pi fC)$. The equivalent circuit for a capacitor is shown in Fig. 43.

FIG. 43.—A capacitor (or condenser) has unwanted inductance in the leads and electrodes and unwanted effective resistance determined by the power losses (page 53). At radio frequencies the inductances and resistances may become very troublesome.

Capacitors (or condensers) are used for many purposes, among which are the following: (*a*) as "blocking" condensers, to prevent the flow of direct current or to impede the flow greatly at low frequencies; (*b*) as "by-pass" condensers which are placed in parallel with devices from which alternating currents are to be excluded and across which negligible alternating voltage is desired (a high-capacitance condenser will have low impedance even at low frequencies, and since $E = IX_C$, if X_C approaches zero, then E_C will approach zero). (*c*) Capacitors are used in conjunction with inductors to form resonant circuits, and (*d*) they are used to cause currents to lead voltages in phase-shifting circuits.

Capacitors for Audio-frequency Circuits.—These are of two general types, those with metal foil for the electrodes and impregnated paper for the dielectric, and the electrolytic type.

Paper condensers consist of two strips of a suitable metal foil insulated with paper and rolled into a bundle. One type is evacuated, and the paper is impregnated with paraffin, wax, or oil, and the assembly is then sealed in an airtight metal can. Another, and very small inexpensive type of paper condenser, is sealed in a waxed cardboard container. The loss in paper condenser is rather high, the power factor being about 0.5 per cent for a typical condenser. However, they are inexpensive, have reasonably high capacitance, and common types will stand direct voltages of the order of 500 volts or more. They are used for blocking and for by-passing alternating currents as previously explained.

Electrolytic condensers are of two types, wet electrolytic condensers and dry electrolytic condensers. The "dry" type is used most widely in radio apparatus, and will be the only type considered here. This condenser is not dry in the strict sense of the word; if it were, it would not function. The so-called "dry" electrolytic condenser consists of aluminum electrodes separated by paper, or gauze, that is saturated with a suitable electrolyte. When a direct voltage is impressed on this condenser, an oxide layer forms on the positive electrode, and this layer is the dielectric. Because the layer is very thin, the capacitance is quite high compared with other condensers of the same physical size. The magnitude of the voltage impressed in forming the dielectric is a factor determing the thickness of the dielectric and hence the working voltage and the capacitance. Because it is necessary to have this forming voltage present at all times when the condenser is in use, an electrolytic condenser must be used in circuits that contain a direct voltage component, such as in power-supply rectifier filter circuits. The leakage of electrolytic condensers is high. They vary in ratings from a few volts and thousands of microfarads (very thin dielectrics) to about 600 volts and 8 microfarads.

Capacitors for Radio-frequency Circuits.—There are two general types, those with gas (including air) for the dielectric and those using solid dielectrics.

Gas (Air) Condensers.—The familiar air condenser is very widely used at radio frequencies, and it is made in a wide variety of voltage ratings and capacitance. In the form used in radio, one set of

fixed plates is the stator, and another set of movable plates is the rotor. Air is a good dielectric under ordinary conditions, and if good insulation is used in an air condenser and if it is otherwise well constructed, the loss is very low. Some high-voltage condensers of this general type are sealed in a suitable container and air or gas at high pressure is injected to serve as a dielectric.

Mica Condensers.—These are widely used at radio frequencies. They are usually of fixed capacitance, but some very small ones are variable. These so-called "trimmers" are about the size of a postage stamp, and the electrode pressure is varied to change the capacitance. High-quality mica is an excellent dielectric, and mica capacitors have low loss. They are usually sealed in a suitable metal or plastic container. Such condensers are made to stand voltages of thousands of volts. Although the capacitance is low (compared with audio-frequency paper and electrolytic condensers) they have ample capacitance for high-frequency radio circuits.

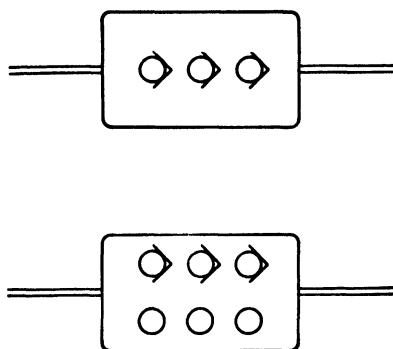



FIG. 44.—The color-coding arrangements for small fixed mica capacitors are as shown.

Ceramic Condensers.—For certain radio-frequency circuits where small stable capacitors having extremely low loss are desired, condensers with metal electrodes and ceramic dielectrics are available. Such dielectrics may contain titanium oxide, which in some forms has a dielectric constant as high as *several thousand* and a power factor that is less than that of mica.

Capacitor Markings.—The capacitance of small radio condensers often is indicated on them by a series of dots that follow the color code of Table VI. If *one row* of three colored dots appears, then the first dot on the left indicates the first significant figure, the middle dot indicates the second significant figure, and the third dot gives the decimal multiplier. For such condensers the voltage rating is 500 volts and the tolerance is 20 per cent. If *two rows* of colored dots are shown (Fig. 44), the top row represents the first three significant figures as read from left to right, and the *bottom row* read from *right to left* gives the decimal multiplier, the

tolerance, and the voltage rating. Capacitances are in micromicrofarads. Thus a 1250-micromicrofarad condenser that has a 2 per cent tolerance and a voltage rating of 1000 volts would be marked brown, red, green, brown, red and gold. Some capacitors are cylindrical, and these are marked with a group of wide and a group of narrow colored bands. With the wide bands *on the right*, the wide bands indicate the significant figure read from *left to right*, and the narrow bands indicate the decimal multiplier, the tolerance, and the voltage rating as read from *right to left*. Attention is called to the fact that this is *not* the only system used and that care must be exercised in the application of Table VI.

TABLE VI.—CONDENSER COLOR CODE[1]

Color	Significant figure	Decimal multiplier	Tolerance, per cent	Voltage rating, volts
Black	0	1		
Brown	1	10	1	100
Red	2	100	2	200
Orange	3	1,000	3	300
Yellow	4	10,000	4	400
Green	5	100,000	5	500
Blue	6	1,000,000	6	600
Violet	7	10,000,000	7	700
Gray	8	100,000,000	8	800
White	9	1,000,000,000	9	900
Gold	....	0.1	5	1000
Silver	....	0.01	10	2000
No color	....	....	20	500

[1] Radio Manufacturers Association, 1938 Standard.

SUMMARY

The frequencies used in radio are of two types, audio frequencies (a-f), from about 50 to 10,000 cycles, and radio frequencies (r-f) that extend from above the audio range to billions of cycles.

The three basic properties of electric circuits are resistance, inductance, and capacitance. Resistance causes a heat loss, and the current through a resistor is in phase with the voltage drop across it. There is no heat loss in a perfect inductor, and the inductance causes the current to lag the voltage drop by 90°. There is no heat loss in a perfect capacitor, and the capacitance causes the current to lead the voltage drop by 90°.

The current flow in a circuit is limited by the impedance, which is composed of the effective resistance and equivalent reactance combined at right angles. That is, from Eq. (7), page 46, $I = E/Z = E/\sqrt{R^2 + X_e^2}$, where X_e is the

equivalent reactance and equals the inductive reactance $X_L = 2\pi fL$ minus the capacitive reactance $X_C = 1/(2\pi fC)$.

If a circuit, or a circuit element, has little loss, then the energy storage factor $Q = X/R$ is said to be high. This term is used extensively in connection with coils, a good coil having a high Q. The reciprocal of the storage factor Q is called the "energy-dissipation factor" $D = R/X$. This term often is used with condensers. Thus a good condenser would have a low dissipation factor.

The magnitude of the resistance of a wire or other conductor of uniform cross section is directly proportional to the length and inversely proportional to the cross-sectional area. For most wires an increase in temperature causes an increase in resistance, as given by Eq. (15), page 51.

The resistors used in radio are of two basic types, wire-wound resistors and composition resistors. Efforts are made to keep the inductance and capacitance low. Wire-wound resistors are used only at audio and the lower radio frequencies.

The inductors used in radio are of two basic types, those having cores of magnetic material and those having air cores. Coils that have laminated silicon-steel cores are used in audio-frequency circuits. Coils that have cores of powdered and compressed magnetic materials are used at audio frequencies and at the lower radio frequencies. Air-cored coils are used at the higher radio frequencies.

The capacitors used in radio are of several types, depending on the dielectric between the electrodes. Paper condensers and electrolytic condensers are used for audio frequencies. Radio-frequency capacitors commonly have dielectrics of air or mica.

REVIEW QUESTIONS

1. Enumerate the types of signals commonly encountered in radio, and give the approximate frequencies for each type of signal.

2. Explain why speech or telegraph signals are not radiated directly into space. What is first done to these signals?

3. What are the basic properties of resistance, inductance, and capacitance?

4. If a voltage is impressed across a coil containing only inductance, or across a condenser containing only capacitance, what limits the current flow?

5. What are the phase relations between the current and voltage in a circuit containing inductance only, and in a circuit containing capacitance only?

6. What term is extensively used in radio to indicate the loss in a coil? To indicate the loss in a condenser?

7. What is meant by the circular mil? By the circular-mil area?

8. In dielectrics at radio frequency what important loss occurs? Describe its nature.

9. Could Eq. (13), page 49, be used to find the volume resistivity of a radio insulator?

10. On page 57, it states that "the remedy for residual inductance may increase the residual capacitance." Explain.

11. Explain how each method shown in Fig. 33 reduces the inductance.

12. What is the difference between a voltage divider and a rheostat?

13. Discuss the types of cores that are used in radio-frequency inductors.

14. Enumerate the types of condensers that are used in audio-frequency circuits and in radio-frequency circuits.

15. On page 68, it states that "mica capacitors have low loss." Would their power factors, storage factors, and dissipation factors be high or low?

PROBLEMS

1. The power-handling capacity of a 1-megohm resistor is 1 watt. What is the maximum current that it can carry and what is the maximum voltage that may be connected across this resistor? Repeat for a ½ watt resistor.

2. Calculate the reactance of a 50-millihenry inductor at 100, 1000, 10,000, 100,000 and 1,000,000 cycles.

3. Calculate the reactance of a 50 micromicrofarad capacitor at the frequencies of Prob. 2.

4. Plot the results of Probs. 2 and 3 on the same set of axes, and determine the frequency at which the reactances of the inductor and capacitor are equal.

5. If the Q of the coil of the preceding problems is 60 at a frequency of 100,000 cycles, what current will flow through a series combination of the inductor and capacitor of Prob. 4 if the voltage is 1.2 volts at the frequency at which the reactances are equal?

6. Measure the dimensions of a typical variable air capacitor adjusted for the maximum setting. Calculate the capacitance and compare with the values given by the manufacturer, or as measured with a bridge.

CHAPTER III

SERIES AND PARALLEL RESONANT CIRCUITS

The preceding chapter was devoted largely to a study of the elements, such as resistors, inductors, and capacitors, that are used to make up audio-frequency and radio-frequency circuits. In introducing such subjects as impedance and reactance, the simple series circuit also was considered.

The circuits of radio transmitting and receiving equipment usually can be broken down into series circuits and parallel circuits. Furthermore, in radio these series and parallel circuits are of the type that are resonant at certain frequencies. One reason for this is that often in radio resonant circuits are used to accept and pass signals of certain frequencies and to reject signals of different frequencies.

Because of the very great importance of series and parallel resonant circuits, much of this chapter will be devoted to their consideration. Other subjects such as measurements also will be included.

Series Circuits.—Electric circuits are composed of resistors, inductors, and capacitors in series, in parallel, or in series-parallel combinations. The theory of a series circuit will be presented by considering typical circuit calculations.

Illustrative Problem.—A coil of 0.06 henry inductance and 8.2 ohms effective resistance, a low-loss condenser of 0.5 microfarad capacitance, and a 25-ohm resistor are all connected in series as shown in Fig. 45 across an audio-frequency alternating voltage of 22.6 volts at 1000 cycles. Make a complete analysis of the circuit.

Solution.—Step 1. Calculate the reactance of the coil and of the condenser, and their equivalent reactance when in series.

$X_L = 2\pi f L = 6.28 \times 1000 \times 0.06 = 377$ ohms.

$X_C = 1/(2\pi f C) = 1/(6.28 \times 1000 \times 0.5 \times 10^{-6}) = -319$ ohms.

$X_e = 377 - 319 = 58$ ohms.

Step 2. Calculate the impedance of the circuit.

The total resistance is $8.2 + 25 = 33.2$ ohms.

From Eq. (7), page 46, $Z = \sqrt{R^2 + X_e^2} = \sqrt{(33.2)^2 + (58)^2} = 66.8$ ohms.

Step 3. Calculate the current from Eq. (7).
$I = E/Z = 22.6/66.8 = 0.339$ ampere or 339 milliamperes.

Step 4. Calculate the power, the power factor and the phase angle, using Eq. (10).
$P = I^2R = (0.339)^2 \times 33.2 = 3.81$ watts.
Power factor $= \cos\theta = P/EI = 3.81/(22.6 \times 0.339) = 0.497 = 49.7$ per cent.
$\theta = \text{arc} \cos 0.497 = 60.2°$.
Power lost in resistor $= I^2R = (0.339)^2 \times 25 = 2.87$ watts.
Power lost in coil $= I^2R = (0.339)^2 \times 8.2 = 0.94$ watt.
Power lost in condenser is assumed to be zero.

Step 5. Compute the voltage drop across each unit, and prove by a vector diagram drawn to scale that the vector sum of the voltage drops equals the impressed voltage.
Voltage across coil $= IZ_L = 0.339 \times \sqrt{(8.2)^2 + (377)^2}$
$= 0.339 \times 378 = 128$ volts.
Voltage across condenser $= IX_C = 0.339 \times 319 = 108$ volts.
Voltage across resistor $= IR = 0.339 \times 25 = 8.5$ volts.

The vector diagram is shown in Fig. 45. In particular it should be noted that very large voltages exist across the coil and condenser.

Algebraic Expression of Vectors.—In the preceding section, three vectors (in this instance the voltage drops across each element of a series circuit) were added to scale graphically to find a resultant vector. Such a graphic solution is not satisfactory in some instances, and a simple means has been devised for expressing vectors algebraically so that these vectors can be added, etc., just like any algebraic quantities.

In this method a vector is broken down by trigonometry into two components, one along the X axis, and one along the Y axis. (Of course other axes of reference can be used.) As an example, suppose that a current of 1 ampere flows through a coil of 3 ohms resistance and 4 ohms reactance, and suppose that the current is assumed to lie along the X axis. There will then be a voltage drop of $IR = 3$ volts *in phase* with the current (along the $+X$ axis) and a voltage drop of $IX = 4$ volts *90° out of phase leading* the current (along the $+Y$ axis, refer to Fig. 30*b*). The vector expression for the total voltage drop (in volts) across the coil would be

$$E_L = 3 + j4, \tag{18}$$

the letter j merely indicating that the value it precedes (in this case the number 4) is at right angles to the axis of reference. The plus sign indicates that inductive reactance is under consideration

and that the total voltage drop of $E_L = \sqrt{(3)^2 + (4)^2} = 5$ volts *leads* the current as it should for a coil. If the resistor and a capacitor had been under consideration, then Eq. (18) would have been written

$$E_C = 3 - j4, \qquad (19)$$

because capacitive reactance is negative and the voltage drop *lags* the current when the current is taken as a reference.

When vector expressions are being combined algebraically and a j^2 term occurs, it immediately is replaced by -1, which it equals by definition. Thus suppose that a current of $I = 3 + j6$ amperes flows through an impedance of $Z = 5 - j7$ ohms, and the vector expression for the voltage drop is desired. This will be

$$\begin{array}{l} 3 + j6 \\ \underline{5 - j7} \\ 15 + j30 \\ \underline{\quad\; - j21 - j^2 42} \\ 15 + j9 \quad - j^2 42 \end{array}$$

and replacing the j^2 by -1 gives $15 + j9 - (-1 \times 42) = 15 + j9 + 42 = 57 + j9$ volts. This method of writing vectors as algebraic expressions will be illustrated further in sections that follow.

Resonance in Series Circuits.—If it happens that the inductive reactance of a series circuit equals

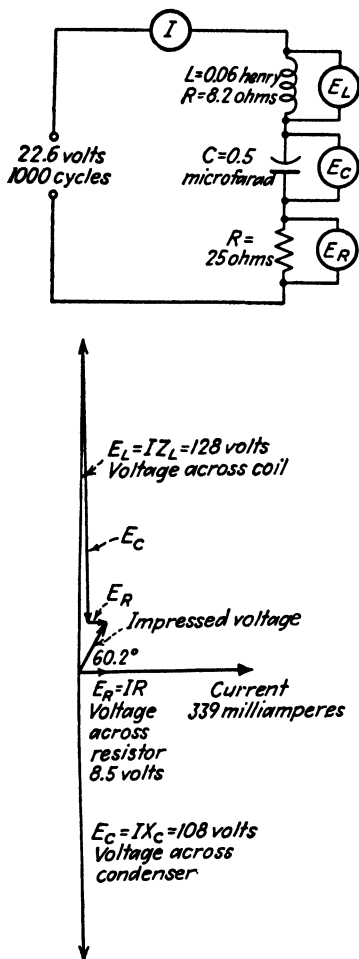



FIG. 45.—When drawing the vector diagram for the series circuit shown, select some suitable scale for the voltages. Plot the voltage across the condenser 90° behind the current (page 45). The voltage across the coil will make an angle θ = arc tan 377/8.2 = 88.7° with the current, and will lead the current (page 44). The voltage across the resistor will be in phase with the current. The resultant or total impressed voltage will be $E_L + E_C + E_R$ as shown, and its position checks the 60.2° angle computed in step 4. The current is taken as the base (along the X axis) in drawing the diagram because the current is the same (or common) to each unit of the series circuit.

the capacitive reactance of that circuit, then the circuit is in **resonance**, or **series resonance** as it is called sometimes. Under these conditions the positive inductive reactance cancels the effect of the negative capacitive reactance, and the impedance of the circuit equals the resistance. At resonance the resistance alone is effective in opposing the flow of current, and the current is in phase with the impressed voltage. This in-phase condition determines when resonance occurs. Such circuits are very important in radio.

Illustrative Problem.—The theory will be illustrated by considering the circuit of Fig. 45 when the impressed voltage is 22.6 volts at 920 cycles, instead of 1000 cycles as previously considered.

Solution.—Step 1. At a frequency of 920 cycles, the reactances are

$$X_L = 2\pi fL = 6.28 \times 920 \times 0.06 = 346.5 \text{ ohms},$$

and

$$X_C = -\frac{1}{2\pi fC} = \frac{1}{6.28 \times 920 \times 0.5 \times 10^{-6}} = -346.5 \text{ ohms}.$$

Step 2. The resistance of the circuit will be the only opposition to current flow, and the current will be $I = E/R = 22.6/33.2 = 0.682$ ampere, instead of 0.339 ampere as it was at 1000 cycles. The voltage drops across each element will be

$$\text{Voltage across coil} = IZ_L = 0.682 \times \sqrt{(8.2)^2 + (346.5)^2} = 0.682 \times 347 = 236.1 \text{ volts}.$$

$$\text{Voltage across condenser} = IX_C = 0.682 \times 346.5 = 236 \text{ volts}.$$

$$\text{Voltage across resistor} = IR = 0.682 \times 32.2 = 22.6 \text{ volts}.$$

A series circuit is in resonance when the inductive reactance equals the capacitive reactance. Equating these two expressions, and solving for the frequency gives

$$X_L = X_C, \quad \text{or} \quad 2\pi fL = \frac{1}{2\pi fC}, \quad \text{and} \quad f = \frac{1}{2\pi\sqrt{LC}}, \tag{20}$$

where f is the frequency of resonance in cycles per second of a *series* circuit, when L is the inductance in henrys, and C is the capacitance in farads. Note that resonance can be obtained by varying either the frequency, the inductance, or the capacitance.

It is noted that at a frequency of 920 cycles the inductive and capacitive reactances are equal, and the resultant *impedance of the circuit at resonance equals the resistance.* If reference is made to the illustrative problem on page 72, it will be observed that the frequency of 1000 cycles is *above* the resonant frequency and that the inductive reactance exceeds the capacitive reactance. The total or resultant impedance is composed of the equivalent reactance $X_e = X_L - X_C$ and the resistance. Since the net or

equivalent reactance would be positive (inductive), the current would lag the voltage (or voltage would lead the current); and the series circuit composed of resistance, inductance, and capacitance would *appear to be inductive* above the resonant frequency, that is, it would be equivalent to a coil having series inductance and resistance. From this reasoning it follows that at any frequency below 920 cycles, the capacitive reactance would exceed the inductive reactance, the net reactance would be negative (capacitive), the current would lead the voltage, and the circuit would appear to be capacitive. This leads to the following general expressions for the impedance of a series circuit containing resistance, inductance and capacitance

$$Z = R + jX_e, \quad \text{or} \quad Z = \sqrt{R^2 + X_e^2}, \quad \text{and} \quad \theta = \text{arc tan}\,\frac{X_e}{R}, \tag{21}$$

where R is the total series resistance and X_e is the equivalent series reactance, $X_e = X_L - X_C$. The characteristics of a series resonant circuit are shown in Fig. 46.

It is now possible to summarize the outstanding features of the series circuit, the last five statements applying *at resonance only*.

a. The current in all parts of a series circuit is the same.

b. The vector sum of all the voltage drops (IX_L, IX_C, and IR) equals the impressed voltage.

c. The equivalent input impedance is the least possible value, and it is pure resistance. It equals in magnitude the resistance of the circuit.

d. The current is the maximum possible value, and it is in phase with the voltage. It equals in magnitude the voltage impressed divided by the resistance of the circuit.

e. The voltage drops across the coil (IX_L) and condenser (IX_C) become quite large with respect to the impressed voltage if the resistance is low and the current rises to a large value.

f. These high voltages may break down condensers or cause other insulation failures, and they are dangerous. On the other hand, the series resonant circuit may be used to increase voltages, because, as seen in this illustration, with 22.6 volts impressed, a *voltage* of 236 volts is available across the condenser or coil.

g. The series circuit has *low equivalent input impedance* and *large input or line current flows* at resonance.

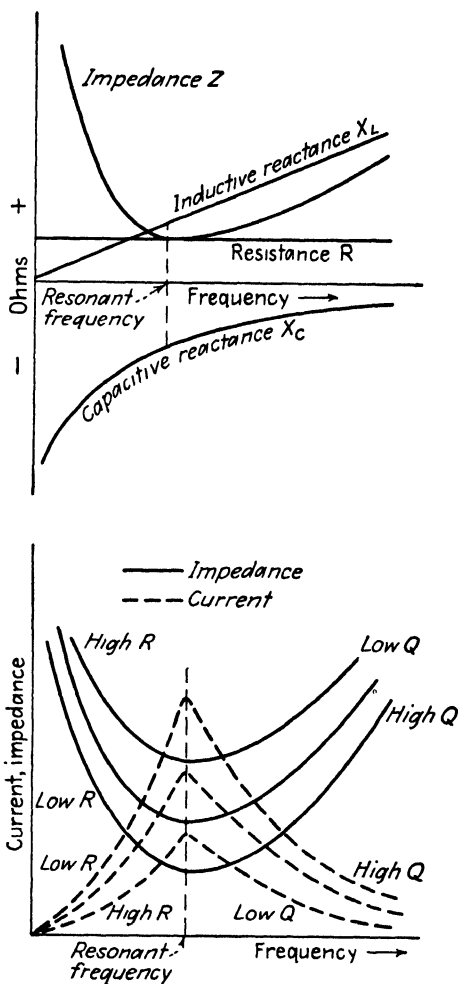
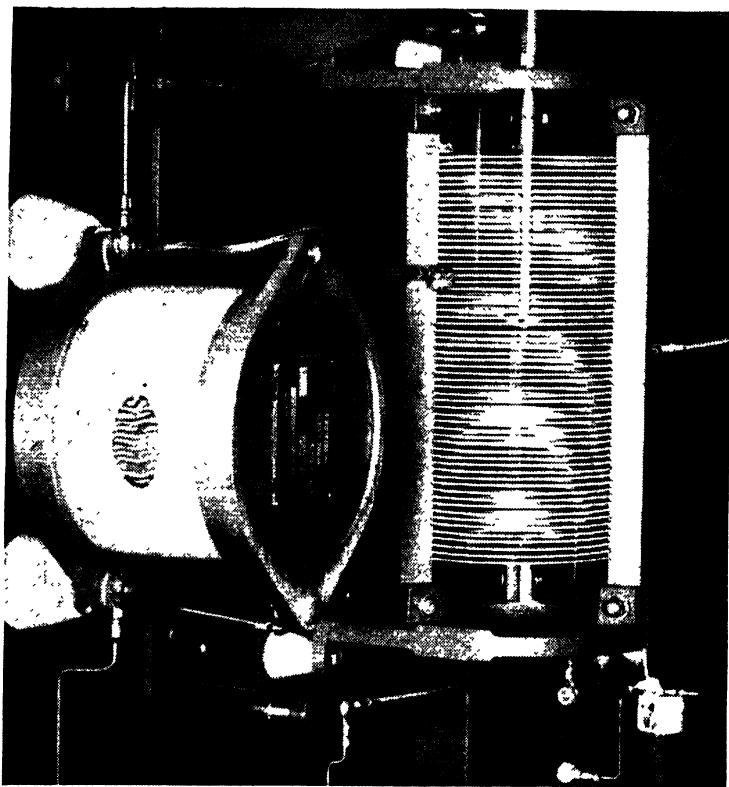



FIG. 46.—The reactances, impedances, and currents of series resonant circuits vary with frequency. At resonance the inductive and capacitive reactances are equal, and the input impedance equals the resistance. The lower figure shows how the resistance and Q influence the magnitudes of the impedances and currents. A low resistance and corresponding high Q give a sharply tuned circuit with a low impedance (equal to the resistance) and high current at resonance.

Parallel Circuits.—These are extensively used in radio. In its most simple form the parallel circuit consists of resistance, inductance, and capacitance all in parallel. In most radio apparatus,

however, the parallel circuit usually consists of an inductor and a capacitor (only) in parallel. The capacitor usually can be considered to have negligible loss, but the effective resistance of the

A tuned resonant parallel circuit consisting of a mica capacitor or condenser (*left*) and an inductor or inductance coil. This parallel circuit serves as a load circuit in the plate of one of the vacuum-tube amplifying stages in a 1000/500 watt amplitude-modulation radio-broadcast transmitter. (*Collins Radio Co.*)

coil must be considered in many instances. The theory of the parallel circuit will be explained by the following example. The solution will be given using vector algebra to illustrate this useful method.[1]

[1] In the illustrations for parallel and series-parallel circuits the resistance of the coil is low, making the computations difficult to perform on a slide rule. Of course it would have been very convenient to have specified a coil with higher resistance. However, this might have been misleading, because coils often have high reactance and low resistance, that is, a high Q. Thus a coil of this type was selected, as giving a more typical illustration.

Illustrative Problem.—Assume that the coil and condenser (only) of Fig. 45 are connected in parallel and across the same voltage source of 22.6 volts at 1000 cycles. Make an analysis of the circuit as shown in Fig. 47.

Solution.—Step 1. Determine the reactance of the condenser and the impedance of the coil. This was done in Step 1 on page 72, where it was shown that $X_C = -319$ ohms. The reactance of the coil was shown to be $X_L = 377$ ohms, and its resistance is 8.2 ohms. The vector expressions for the impedances of the condenser and coil are

$$Z_C = 0 - j319 \text{ ohms} \qquad \text{and} \qquad Z_L = 8.2 + j377 \text{ ohms.}$$

Step 2. Calculate the current that will flow through each of these parallel units. In these solutions the voltage will be taken as the base, because it is common to both the condenser and the coil. Thus the vector expression for the voltage will be $E = 22.6 + j0$.

$$I_C = \frac{E}{Z_C} = \frac{22.6 + j0}{0 - j319} = \frac{0 + j7220}{101{,}800 + j0} = 0 + j0.0708 = 0.0708 \text{ ampere.}$$

$$I_L = \frac{E}{Z_L} \times \frac{22.6 + j0}{8.2 + j377} = \frac{185 - j8530}{142{,}067 + j0} = 0.0013 - j0.06$$
$$= 0.06 \text{ ampere (approximately).}$$

For review purposes, the details of the last calculation are shown here. In dividing one vector expression by another, they are each multiplied by the "opposite" (more correctly, the conjugate) of the divisor to eliminate the j term below the division sign. When j^2 occurs it is replaced with -1.

$$\frac{22.6 + j0}{8.2 + j377} \times \frac{8.2 - j377}{8.2 - j377} = \frac{185 - j8530}{142{,}067} = 0.0013 - j0.06 \text{ ampere}$$

$$\begin{array}{l} 22.6 + j0 \\ \underline{8.2 - j377} \\ 185 + j0 \\ \underline{\quad - j8530 - j^2 0} \\ 185 - j8530 \end{array} \qquad\qquad \begin{array}{l} 8.2 + j377 \\ \underline{8.2 - j377} \\ 67 + j3090 \\ \underline{\quad - j3090 - j^2 14{,}200} \\ 142{,}067 + j0 \end{array}$$

Step 3. Calculate the total current supplied by the source of voltage. From Step 2 it is seen that the capacitor current $I_C = 0 + j0.0708$ ampere is a current that *leads* the voltage (which was taken as the base) by 90°. It is also seen that inductor current $I_L = 0.0013 - j0.06$ ampere is a current that *lags* the voltage by almost 90°. The current I_T supplied by the source will be the vector sum of these.

$I_T = (0 + j0.0708) + (0.0013 - j0.06) = 0.0013 + j0.0108 = 0.0109$ ampere (approximately). It will be observed that the resultant or total current is smaller than either of the branch currents and that it leads the voltage by the angle θ = arc tan $0.0108/0.0013 = 8.32$, and $\theta = 83.15°$.

Step 4. Calculate the equivalent impedance of the parallel combination. This will be the impressed voltage divided by the total or resultant current.

$$Z = \frac{E}{I} = \frac{22.6 + j0}{0.0013 \quad j0.0108} = \frac{0.0294 - j0.244}{0.0001184} = 248 - j2060$$
$$= 2075 \text{ ohms.}$$

As a check, $Z = E/I = 22.6/0.0109 = 2075$ ohms (approximately).

Resonance in Parallel Circuits.—In the illustrative problem of the preceding section, at a frequency of 1000 cycles the lagging current taken by the coil almost completely neutralized the leading current taken by the condenser. Because of this the resultant line current flowing from the voltage source to the parallel circuit was small. This small current was, however, leading, and from this it follows that the input impedance above resonance for a parallel circuit is capacitive. If the frequency of the voltage impressed on a parallel circuit is of a value such that the reactive component (90° out of phase) of the current through the coil equals the reactive component of the current through the condenser, then the reactive components would completely cancel each other. Under these conditions, the only current flowing from the source to the parallel combination would be a small in-phase component flowing to supply the power consumed by the resistance of the coil. Thus at resonance the input impedance of a parallel circuit would be a large value of pure resistance. For the circuit of Fig. 47, this condition will be approximated at a frequency of 918 cycles. Below the resonant frequency the out-of-phase component of the current taken by the coil will exceed the reactive current taken by the condenser, and the input impedance of the parallel circuit will be inductive.

Fig. 47.—When a coil and a condenser are in parallel, the total line current is the vector sum of the branch currents. In drawing the vector diagram, the voltage is taken as the base because it is common to each branch. The current through the condenser is 90° ahead of the voltage (page 45), and the current through the coil is approximately 90° behind the voltage (page 44). The line current I_T is very small in a sharply tuned circuit (one with low-loss condensers and coils), and the input impedance is a high value.

Illustrative Problem.—As an example of resonance in parallel circuits, the illustrative problem of the preceding section will be analyzed, but with a voltage of 22.6 volts at 918 cycles.

Solution.—Step 1. The reactances will be

$$X_C = \frac{1}{2\pi f C} = \frac{1}{6.28 \times 918 \times 0.5 \times 10^{-6}} = -346.7 \text{ ohms}$$

and

$$X_L = 2\pi f L = 6.28 \times 918 \times 0.06 = 346.08 \text{ or } 346.1 \text{ ohms.}$$

The corresponding impedances will be

$$Z_C = 0 - j346.7 = 346.7 \text{ ohms}$$

and

$$Z_L = 8.2 + j346.1 = \sqrt{(8.2)^2 + (346.1)^2} = 346.2 \text{ ohms (approximately).}$$

Step 2. The current through the condenser will be

$$I_C = \frac{22.6}{346.7} = 0.0652 \text{ ampere,}$$

and the current through the coil will be

$$I_L = \frac{22.6}{346.2} = 0.0653 \text{ ampere.}$$

The current through the condenser will *lead* the voltage by 90°, and the current through the coil will lag the voltage by the angle θ = arc tan 346.1/8.2 = 42.2, and θ = 88.65° (approximately).

Step 3. The current of 0.0653 ampere through the coil will consist of two components, one I_X = 0.0653 sin 88.65 = 0.0653 × 0.9997 = 0.0652 ampere, which is 90° out of phase with the voltage, and the other I_R = 0.0653 cos 88.65° = 0.0653 × 0.02356 = 0.00154 ampere, which is in phase with the voltage.

This is the current that at resonance will flow from the generator to the parallel circuit, and for a given impressed voltage this in-phase current is determined by the resistance of the coil. The equivalent impedance offered by the parallel circuit is the voltage impressed divided by the current that flows, or

$$Z = \frac{E}{I} = \frac{22.6}{0.00154} = 14{,}650 \text{ ohms.}$$

This will be pure resistance because the line current flowing from the generator to the parallel circuit is in phase with the impressed voltage.

A parallel circuit is in resonance when the out-of-phase component of the current through the coil equals the current through the condenser (a low-loss condenser in which the current leads the voltage by 90° being assumed). The current through the condenser is $I_C = E/X_C$. The out-of-phase component of the current through the coil is $I_L = (E/Z_L) \sin\theta = (E/Z_L)(X_L/Z_L)$. Equating these two quantities to find the resonant frequency,

$$\frac{E}{X_C} = \frac{EX_L}{Z_L^2}, \quad \text{or} \quad \frac{1}{X_C} = \frac{X_L}{Z_L^2}, \quad \text{and}$$

$$f = \frac{1}{2\pi}\sqrt{\frac{L - CR_L^2}{CL^2}}. \qquad (22)$$

This equation gives the frequency f in cycles per second at which the line current flowing from the generator to the parallel load is in phase with the line voltage. The term C is the capacitance in farads of the condenser, assumed to be lossless. The term L is the inductance in henrys, and R_L is the resistance of the coil in ohms. Note that if R_L equals zero, Eq. (22) becomes the same as Eq. (20). Thus it can be said that in a parallel circuit the resistance affects the resonant frequency of the circuit; but for most radio circuits where R_L is low (the Q is high) this effect is slight, and Eq. (20) is satisfactory for both series and parallel circuits. In any event, radio circuits usually are adjustable and they generally are tuned to resonance experimentally; thus Eq. (20) usually is satisfactory. Another point is that when a coil has appreciable resistance the condition of unity power factor is not the same as the condition of maximum impedance, and this latter may be the condition of operation desired. But again, Eq. (20) usually is of sufficient accuracy, and is used extensively in radio.

The general expression for the input impedance of a parallel circuit *at any frequency* can be determined as follows: When a source of voltage of E volts is connected across a parallel circuit, a total or line current $I_T = E/Z_P$ will flow, where Z_P is the input impedance of the parallel circuit. As has been seen in the preceding pages, a current $I_C = E/Z_C$ will flow through the condenser, and a current $I_L = E/Z_L$ will flow in the coil. Now the total or line current will be the vector sum of these individual currents, thus

$$I_T = I_C + I_L, \quad \text{or} \quad \frac{E}{Z_P} = \frac{E}{Z_C} + \frac{E}{Z_L}, \quad \text{and}$$
$$\frac{1}{Z_P} = \frac{1}{Z_C} + \frac{1}{Z_L}, \quad \text{whence} \quad Z_P = \frac{Z_C Z_L}{Z_C + Z_L}. \tag{23}$$

An example of the use of this equation will be given in the next section. The characteristics of a parallel resonant circuit are shown in Fig. 48.

Radio circuits usually are composed of coils and condensers that have low losses and therefore low equivalent series resistance. Also, in such circuits the reactances encountered quite often are large compared with the resistances. For this reason, at resonance Eq. (23) can be simplified by neglecting the resistance components *in the numerator only,* by assuming that the capacitive reactance

$X_C = 1/(2\pi fC)$ equals the inductive reactance $X_L = 2\pi fL$, and by letting the combined series impedance of Eq. (23) be $Z_C + Z_L = R_L$. This assumes that the reactances of the denominator

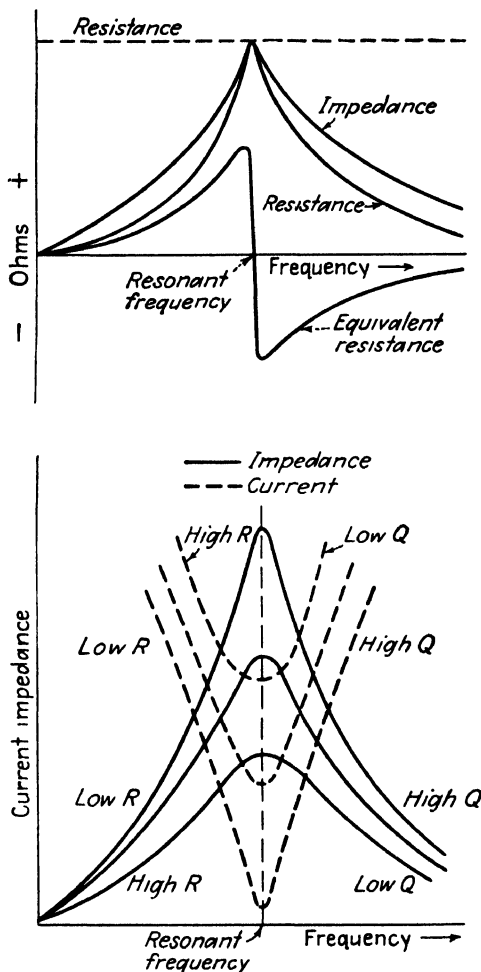



FIG. 48.—The input resistance, equivalent reactance, and impedance of a parallel resonant circuit vary with frequency. The lower figure shows how the resistance and Q influence the input or equivalent impedance and the current. A low resistance and a high Q give a sharply tuned circuit with high input impedance and low input current at resonance.

cancel, and that the series resistance of the condenser is zero. When these approximations are made, the input impedance at resonance equals *pure resistance* of magnitude

$$Z_P = R_P = \frac{X_C X_L}{R_L} = \frac{X_L^2}{R_L} = \frac{(\omega L)^2}{R_L} = \omega LQ, \qquad (24)$$

where Q equals $\omega L/R_L$, and is the Q of the coil. This equation is for *resonance only* and holds quite closely only for circuits having condensers with negligible losses and coils with a high Q, in the vicinity of 100.

It is now possible to summarize the outstanding features of a parallel circuit, the last five statements applying *at resonance only.*

a. The voltage across the different branches of a parallel circuit is the same.

b. The vector sum of the branch currents equals the line current.

c. The input impedance rises to a high value, and is pure resistance. It is determined, in part, by the resistance of the parallel branches, but is *greatly* different from the branch resistances in most radio circuits.

d. The current flowing to the parallel circuit falls to a very low value, and is in phase with the voltage. The magnitude is $I = E/Z_P$ or $I = E/(\omega LQ)$ (approximately).

e. The current in the condenser (E/Z_C) and the current in the coil (E/Z_L) may be quite large, exceeding the total or line current many times.

f. These high currents which may flow in the branches are not dangerous, but the high voltages which may exist in a series circuit at resonance are dangerous. The high current in the coil is very useful in inducing large voltages *at the resonant frequency* in adjacent coils, a principle that is used often in radio.

g. The parallel circuit has *high input impedance* and *low input or line current flow* at resonance.

It will be noted that series circuits and parallel circuits act oppositely to each other. For this reason the phenomena described in the preceding sections sometimes are called "resonance" (for the series circuit) and sometimes "antiresonance" (for the parallel circuit). They also are called "voltage resonance" (series circuit) and "current resonance" (parallel circuit).

Series-parallel Circuits.—Some circuits contain series combinations and parallel combinations connected across the same source of voltage, Fig. 49 being an example. To solve such a circuit it is first necessary to change the parallel circuit to an equivalent series circuit. The solution will be given now, using polar coordinates to express the vectors. In the polar system the magnitude and

the angle of a vector are specified, instead of stating the two components and preceding the reactive term by the letter j.

Illustrative Problem.—Solve the circuit of Fig. 49 for the current through the capacitor and the current through the inductor.

Solution.—Step 1. Find the impedance of the branches of the parallel portion. This was done on page 81, where it was shown that

$Z_C = 0 - j346.7 = 346.7 \angle -90°$ ohms and
$Z_L = 8.2 + j346.1 = 346.2 \angle +88.65°$ ohms,

the angles being found by the method indicated on page 48.

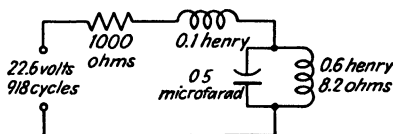



FIG. 49.—A series-parallel circuit.

Step 2. Calculate the equivalent impedance of the parallel portion, using Eq. (23).

$$Z_P = \frac{Z_C Z_L}{Z_C + Z_L} = \frac{346.7 \angle -90° \times 346.2 \angle +88.65°}{(0 - j346.7) + (8.2 + j346.1)} = \frac{119{,}800 \angle -1.35°}{8.2 - j0.6} =$$

$$\frac{119{,}800 \angle -1.35°}{8.25 \angle -4.19°} = 14{,}550 \angle +2.84° \text{ ohms.}$$

This agrees reasonably well with the value for the equivalent impedance obtained on page 81.

Step 3. Calculate the impedance of the series portion of the circuit.

$Z_S = 1000 + j(6.28 \times 918 \times 0.1) = 1000 + j527 = 1130 \angle +27.8°$ ohms.

Step 4. Add the series impedance to the equivalent parallel impedance to determine the total impedance connected to the source.

$Z_P = 14{,}550 \angle +2.84° = 14{,}550 \cos 2.84° + j14{,}550 \sin 2.84° =$
$14{,}550 \times 0.9987 + j14{,}550 \times 0.0494 = 14{,}300 + j720$ ohms.
$Z_T = (14{,}300 + j720) + (1000 + j527) = 15{,}300 + j1247 =$
$15{,}300 \angle +4.65°$ ohms (approximately).

Step 5. Assume the impressed voltage as the reference, and calculate the current that will flow from the generator.

$$I = \frac{E}{Z_T} = \frac{22.6 \angle 0°}{15{,}300 \angle +4.65°} = 0.001475 \angle -4.65° \text{ ampere.}$$

That is, the generator current will be 1.475 milliamperes, and will lag the generator voltage by 4.65°.

Step 6. Calculate the voltage drop caused by the current flowing in the series portion of the circuit.

$E_S = IZ_S = (0.001475 \angle -4.65°) \times (1130 \angle +27.8°) = 1.67 \angle +23.15°$
$= (1.67 \cos 23.15°) + j(1.67 \sin 23.15) = 1.53 + j0.655$ volts.

Step 7. Find the voltage across the parallel portion by subtracting the voltage drop across the series part from the impressed voltage.

$$E_P = (22.6 + j0) - (1.53 + j0.655) = 21.07 - j0.655 = 21.1 \angle{-1.78^\circ} \text{ volts.}$$

Step 8. Calculate the current in each branch of the parallel circuit.

$$I_C = \frac{21.1 \angle{-1.78^\circ}}{346.7 \angle{-90^\circ}} = 0.0609 \angle{+88.22^\circ} \text{ ampere.}$$

$$I_L = \frac{21.1 \angle{-1.78^\circ}}{346.2 \angle{88.65^\circ}} = 0.061 \angle{-90.43^\circ} \text{ ampere.}$$

Thus the current through the condenser is about 61 milliamperes and leads the source voltage by about 88°, and the current through the coil is approximately 61 milliamperes and lags the source voltage by about 90°.

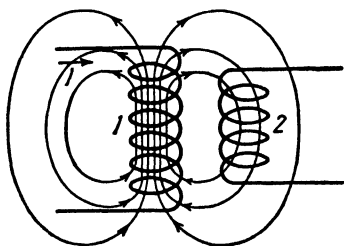



FIG. 50.—The lines of magnetic force or flux produced by coil 1 link with coil 2 and will induce a voltage in coil 2 if the flux linkages are changed. These two coils are assumed in air. If the two coils are on a continuous iron core, almost all the flux produced by coil 1 will link coil 2.

Mutual Inductance.—If two circuits are situated as in Fig. 50 so that the current in coil 1 produces a magnetic field that links coil 2, then mutual inductance exists between the two coils. If the current in coil 1 changes, then the magnetic field linking with the turns of coil 2 also changes, and this change of flux linkages induces a voltage in coil 2. *Mutual inductance is the property of two circuits that causes a voltage to be induced in one circuit when the current in the other circuit is changed.* Transformers and similar devices operate because of mutual inductance between the windings. Mutual inductance is measured in **henrys.**

The magnitude of the voltage induced in one coil (the secondary) by an alternating current in another coil (the primary) is given by the relation

$$E_S = 2\pi f M I_P, \tag{25}$$

where E_S is the induced secondary voltage in volts, f is the frequency in cycles per second of the primary current, M is the mutual inductance in henrys between the primary and secondary, and I_P is the alternating current in amperes in the primary. The use of this equation will be explained in the following chapter (page 102).

It is of interest to note that a certain amount of voltage is induced in series in *each* turn of the secondary and that each of these series-induced voltages adds to produce the voltage between the ends of the coil. Thus the voltage induced in the secondary is a *series* voltage, and this produces a voltage between the ends of the coil. A perfect transformer which has negligible losses and which is wound in the usual way produces a 180° shift in phase between the primary and secondary voltages. If Fig. 50 is a lossless transformer the induced, or back, voltage in the primary must be 180° out of phase with the voltage impressed on the primary. The same magnetic flux that induces the back voltage in the primary also induces the secondary voltage. Thus if at a given instant the voltage of the *upper end* of the primary (coil 1 of Fig. 50) is positive, at that same instant the *lower end* of the secondary (coil 2) will be positive.

Nonsinusoidal Waves.—The theory and calculations of the preceding pages have been based on sinusoidal, or sine-wave, voltages and currents. This is common practice. Also, the values of voltage and current specified were effective values, that is, the values which would be indicated on the usual types of voltmeters or ammeters when connected in the circuits. The question arises: Why are calculations made on the basis of sinusoidal voltages and currents when it is well known that speech signals and telegraphic code signals are not sinusoidal?

The answer to this question is involved, and it cannot be given in detail at this time. Suffice it to say, for the present, that experience has shown that if a circuit is designed to pass sine-wave signals over the required frequency band without excessive distortion, then the circuit also will pass nonsinusoidal signals without excessive distortion.

Nonsinusoidal Nonrecurrent Waves.—Certain signals such as speech are nonsinusoidal and nonrecurrent, in the sense that each portion of a speech signal is different, unless of course if the same word was being spoken over and over again in the same way. However, as previously mentioned, if circuits and equipment are designed on a sine-wave basis to pass satisfactorily a sufficiently wide frequency band, then speech or other similar signals will be transmitted with a high degree of fidelity.

Nonsinusoidal Recurrent Waves.—Certain waves, such as the output of a rectifier or of a square-wave oscillator are nonsinusoidal,

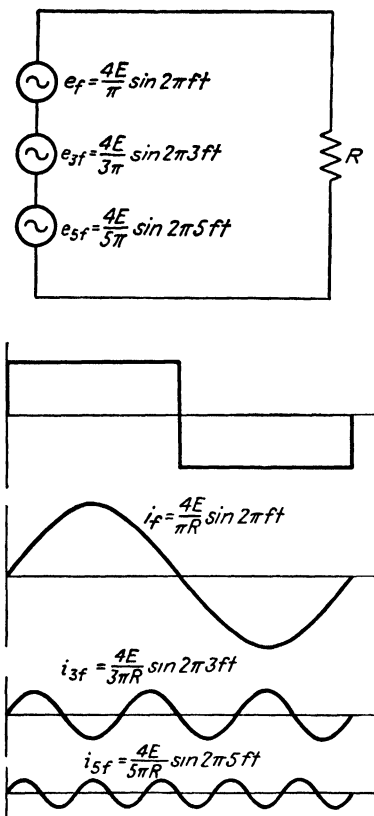



FIG. 51.—If three generators e_f, e_{3f}, and e_{5f} having the voltages indicated and being in phase are connected to a resistor R, each generator will force a corresponding current i_f, i_{3f}, and i_{5f} through the resistor. These separate voltages, or these separate currents, combine to give (approximately) the rectangular wave form shown. The equation for a rectangular voltage wave is

$$e = \frac{4E}{\pi}\left(\sin 2\pi ft + \frac{1}{3}\sin 2\pi 3ft + \frac{1}{5}\sin 2\pi 5ft + \ldots\right),$$

where E is the maximum value of the rectangular wave, and f is its frequency. Harmonic components higher than the fifth have been neglected.

but are recurrent in that, once established, they occur time after time in the same way.

If a nonsinusoidal wave, no matter how complicated the wave form, occurs the same, time after time, then that wave can be shown to be composed of a sinusoidal fundamental and a series of sinusoidal components or harmonics. This point of view will be explained now. As shown in Fig. 51, several sine-wave generators, each having the voltage of magnitude and phase as indicated, are connected in series to a resistor. Each generator acts individually to force a sine-wave current through the resistor. These individual sine-wave currents add together to produce what approximates a square-wave current through the resistor and a square-wave voltage across the resistor. Thus if the various sinusoidal components of Fig. 51 combine to give a square wave, then it can be stated that a square wave is composed of a series of sine waves each of the proper magnitude, frequency, and phase relation. A series of telegraph dots or dashes consist of a large number of sinusoidal components much as in Fig. 51.

A saw-tooth wave or any other nonsinusoidal recurrent wave can be built up from sinusoidal waves in a similar manner (see also page 408).

Symmetrical and Nonsymmetrical Waves.—The square wave of

Fig. 51 was symmetrical about the zero or X axis. It will be noted that this wave contained only *odd* harmonics. Waves containing only odd harmonics always are symmetrical, that is, the positive and the negative half cycles always are *exactly* the same. Waves which contain *even* harmonics are nonsymmetrical, that is, the positive and negative half cycles are different. As will be shown in Chap. VII, a rectified wave contains even harmonics, and it is nonsymmetrical; in fact, the negative half cycle is zero.

Measuring Instruments.—Measurements in audio-frequency and radio-frequency circuits are difficult for several reasons: (*a*) The power level in the circuit is low, and the measuring instrument must absorb negligible power; (*b*) the frequency range is very wide and certain types of instruments, satisfactory at low frequencies, are unsatisfactory for communication purposes; (*c*) in high-frequency radio circuits the presence of the measuring circuit may unbalance normal conditions, giving misleading values.

Fig. 52.—Connections of the elements of a copper oxide measuring instrument.

Progress in the field of measurements has been very rapid in recent years, and many excellent instruments are available. The basic types will be considered now, leaving those containing vacuum tubes for discussion in later chapters.

Copper Oxide Rectifier Instruments.—If a thin layer of cuprous oxide (usually called "copper oxide") is formed on a layer of copper, electron current readily flows from the copper to the oxide, but flows with difficulty in the reverse direction. Stacks of copper oxide disks are used to rectify alternating waves, and the rectified direct-current component is measured with an ordinary direct-current instrument.

A common arrangement is shown in Fig. 52, in which the rectifier elements indicate the direction in which current flows readily. If this circuit is studied, it will be evident that for either polarity of the applied alternating signal, current will flow *down* through the direct-current indicating instrument.

The circuit of Fig. 52 can be used as an ammeter; or a high

protective resistance can be placed in series with the circuit, and the combination used as a voltmeter. Such devices are used extensively as voltmeters, and to some extent as ammeters. The accuracy of these instruments begins to decrease at about 5000 cycles, and they are not very satisfactory above about 10,000 cycles. One important reason for this is the capacitance between the disks. Of course they can be calibrated for higher frequencies, but in general they are used only over the audio-frequency range.

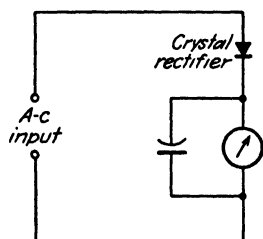



Fig. 53.—Connections for a crystal rectifier. The condenser is to provide a path for the alternating-current components. The rectified direct-current component causes the instrument to deflect.

Crystal Rectifiers.—Early in the development of radio it was found that if a fine wire, perhaps of tungsten (and often reffered to as a "whisker" or a "cat's whisker") firmly touched one of the various crystals (such as silicon or germanium) the contact had rectifying properties. Such devices were used extensively and then decreased in importance, but are once again widely used, particularly at ultrahigh frequencies.

A circuit is shown in Fig. 53. The alternating signal to be measured is impressed on the crystal as indicated. A rectified current flows through the crystal rectifier, and the direct-current component causes a deflection of the indicating instrument. The

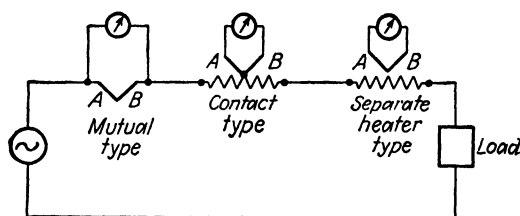



Fig. 54.—Three methods used for heating a thermocouple junction. Other schemes are used. In these illustrations the heater has been shown as a resistor to stress the principle of operation. Often the heater is drawn as a heavy straight line.

condenser provides a low-impedance path for alternating-current components, thus permitting almost the entire signal input to appear across the crystal, which gives a more sensitive device. Sometimes the condenser is omitted, sufficient capacitive effect being supplied by the connecting leads.

Thermocouples.—When the junction of two dissimilar wires is heated, a direct voltage exists between the ends of the wires. If the ends of the wires are connected to a sensitive direct-current instrument, a current will flow and the pointer will deflect. Three methods of heating the thermocouple junction are shown in Fig. 54. In the mutual type the alternating current to be measured passes through the two dissimilar wires *A-B*. Because of their resistance the wires are heated, a direct voltage is produced, and the indicating instrument deflects. In the contact type the alternating current to be measured passes through a fine resistance wire and causes the heating effect. The separate-heater type is similar except that the thermal junction is not in electrical contact with the heater.

FIG 55—The delicate wires of a thermocouple often are mounted in an evacuated glass bulb, and this is placed in a suitable holder. (*General Radio Co*)

Thermocouples of the separate-heater type are designed especially for radio measurements. A bead of some electrical-insulating and heat-conducting material separates electrically the thermocouple junction, the connecting leads, and the indicating instrument from the radio-frequency circuit in which measurements are to be made. This prevents, in part, errors in measurement. A specially designed thermocouple of this type is very accurate up to hundreds of megacycles. Such thermocouples are mounted in an evacuated bulb, as shown in Fig. 55, to protect the delicate wires and to reduce the heat loss.

The effective value of an alternating current is defined as the value that causes the same heating effect in a given resistor as a given value of direct current. For this reason, thermocouples can be calibrated by connecting the heater in series with a resistor, a direct-current milliammeter, and a battery. A curve plotted with deflection against current, as read on the direct-current instrument, also will apply when the thermocouple is used in alternating-cur-

rent circuits. A thermocouple also can be used with an adequate series-protecting resistor to measure alternating voltages. A thermocouple draws power from the circuit in which it is used, and this limits its application in some instances. Certain types of thermocouple instruments cannot be calibrated by the simple direct-current arrangement discussed, and must be calibrated on alternating current.

SUMMARY

In a series circuit at resonance the inductive reactance equals the capacitive reactance, the current rises to a large value, and the input impedance falls to a low value, a large voltage drop may exist across the coil and the condenser, and the current through the circuit is in phase with the voltage across the circuit. At resonance the input impedance is a low value of pure resistance. The frequency of resonance is given by Eq. (20) to be $f = 1/(2\pi\sqrt{LC})$. A detailed listing of the characteristics of a series circuit at resonance is given on page 76.

In a parallel circuit at resonance the out-of-phase component of the current through the coil equals the current through the (perfect) condenser, the line current falls to a low value, and thus the input impedance rises to a large value, and the current taken by the parallel circuit is in phase with the voltage across the circuit. At resonance the input impedance is a high value of pure resistance. The frequency of resonance is given, approximately, by the equation of the preceding paragraph. The equivalent impedance of any parallel circuit is given by Eq. (23) [for low-loss circuits at resonance by Eq. (24)]. A detailed listing of the characteristics of a parallel circuit at resonance is given on page 84.

It is of much importance to note that series and parallel circuits act oppositely. For this reason the terms "resonant circuit" (series) and "antiresonant circuit" (parallel) often are used.

Through the effect of mutual inductance, a voltage E_S is induced in the secondary when the current I_P in the primary is changed. The magnitude of the induced voltage is $E_S = 2\pi f M I_P$.

Although speech waves and telegraphic signals are nonsinusoidal, circuits properly designed on a sine-wave basis will pass such complex signal waves. When nonsinusoidal waves are recurrent, they are composed of a fundamental and a series of harmonics. Nonsinusoidal symmetrical waves contain only odd harmonics, and nonsinusoidal nonsymmetrical waves contain even harmonics.

Copper oxide instruments are very useful for measurements at audio frequencies, but at radio frequencies crystals, thermocouples, or special vacuum-tube instruments must be used.

REVIEW QUESTIONS

1. State one reason that series and parallel resonant circuits are used extensively in radio apparatus.

2. State the two fundamental principles of a series circuit.

3. If j^2 occurs in a computation, what is done?

4. In a series resonant circuit the voltages across the coil and condenser may be many times greater than the source voltage. What will happen if the resistance of the circuit is decreased to a very low value?

5. Name three ways in which the resonant frequency of a series circuit can be changed.

6. In Step 3 of the illustrative example on page 79 it states that the inductor current lags by almost 90°. Why is not the angle exactly 90°?

7. Is resonance in a parallel circuit obtained when the capacitive and inductive reactances are equal? Explain.

8. Why is it possible in many problems to assume that the phase angle of a condenser is 90°? For what types of condensers would this assumption be in considerable error?

9. Briefly compare the characteristics of series and parallel circuits at resonance.

10. What is the nature of the input impedance of a series circuit and of a parallel circuit below resonance, at resonance, and above resonance?

11. Will a series circuit composed of a coil and condenser be resonant at exactly the same frequency as a parallel circuit composed of the same coil and condenser?

12. Give a common example of a nonsinusoidal, nonsymmetrical wave. What important frequency components must it contain.

13. On page 90 the statement is made that the capacitance between disks limits the radio-frequency use of the copper oxide rectifier. Why?

14. Refer to Fig. 53, page 90, and explain why the condenser is used. Must it always be used?

15. On page 92 the statement is made that in some instances the application of a thermocouple is limited by the power drawn. Explain why this is true.

PROBLEMS

1. Repeat the calculations of the illustrative problem of page 75 at a frequency of 850 cycles, and compare your results with the discussion on page 76 immediately preceding Eq. (21).

2. Repeat the calculations of the illustrative problem of page 79 at a frequency of 850 cycles, and compare your results with the computations made in the book at 1000 cycles, and at 918 cycles.

3. Repeat the calculations of the illustrative problem of page 85 if the series coil has an inductance of 0.2 henry.

4. A voltage of 1.0 volt at 1,000,000 cycles is impressed across a coil that has negligible resistance and 20 microhenrys inductance. What current will flow in the coil? If a second coil is brought near the first coil and a voltage of 0.1 volt is induced in it, what is the mutual inductance between the two coils?

5. The heater of a sensitive thermocouple has a resistance of 1000 ohms, and will carry a maximum current of 1.8 milliamperes. A protective resistor of 12,000 ohms is placed in series with the thermocouple to measure a voltage.

When the current through the thermocouple heater is 1.7 milliamperes, as determined by the calibrated direct-current indicating instrument, what is the magnitude of the unknown voltage? What is the total power taken from the circuit under test? What power is dissipated in the protective resistor and in the thermocouple heater?

CHAPTER IV

POWER TRANSFER AND IMPEDANCE MATCHING

The preceding chapters reviewed those electrical principles of particular importance in radio, and also discussed the circuit elements, such as resistors, inductors, and capacitors, used in radio circuits. In addition, series, parallel, and series-parallel combinations of these elements were discussed, and a brief survey of measuring instruments was given. Also, the very important subjects of series and parallel resonance were considered.

Radio transmitting and receiving sets are composed of various units, or sections, that accomplish certain operations. These sections are connected together and electric signals pass from one section to another. Often this involves the transfer of electric power, and the power transfer depends on the impedance match between the sections. Because these impedances may not be the best values, impedance changes or transformations may be necessary. This chapter will discuss these, and related, subjects, and will present the principles of operation of circuits that are used for transforming and matching impedances.

The Decibel.[1]—This unit is used in radio, and in other branches of communication, to measure the decrease in signal strength when a signal passes through a circuit containing attenuation (loss), and to measure the increase in signal strength when a signal passes through a circuit containing amplification (gain). The decibel (abbreviated db) is used primarily to measure or express *ratios* of power, voltage, and current. It also is used in radio to measure the field strength of radio signals from antennas and to make acoustical measurements (page 14).

For power ratios the decibel is used as follows: If the electric power is flowing into a circuit or device, such as a pad or attenuator, that *decreases* the signal strength, then the loss caused by the device is

[1] For a more complete treatment of this subject, consult A. L. Albert, "Electrical Fundamentals of Communication," McGraw-Hill Book Company, Inc.

$$n = 10 \log_{10} \frac{P_1}{P_2}, \tag{26}$$

where P_1 is the power input (in kilowatts, watts, milliwatts, etc.), P_2 is the power output in corresponding units, and n is the *power* loss in decibels. If electric power is flowing into a device, such as certain types of amplifiers, that *increases* the signal strength, then the power gain produced by the device is

$$n = 10 \log_{10} \frac{P_2}{P_1}, \tag{27}$$

where P_2 is the power output, P_1 is the power input measured in the same units, and n is the *power* gain in decibels.

For voltage ratios the decibel is used as follows: If a signal voltage is impressed on a circuit or device that *decreases* the voltage, then the voltage loss in decibels caused by the device is

$$n = 20 \log_{10} \frac{E_1}{E_2}, \tag{28}$$

where E_1 is the impressed voltage, E_2 is the output voltage, and n is the *voltage* loss in decibels. If the circuit or device is an amplifier that *increases* the voltage, then the voltage gain is

$$n = 20 \log_{10} \frac{E_2}{E_1}, \tag{29}$$

where E_2 is the voltage output, E_1 is the voltage input, and n is the *voltage* gain in decibels.

For current ratios the decibel is used as in Eqs. (28) and (29) with the currents I_1 and I_2 being substituted for the voltages E_1 and E_2. It should be emphasized that in a given circuit voltage ratios and current ratios give the same number of decibels as power ratios only under special conditions.

In communication circuits it is difficult to measure electric power directly with instruments because the amounts of power are low and the frequency is high. Because of this, *power* losses in decibels often are determined from voltage measurements (and ratios) using, perhaps, a vacuum-tube voltmeter, or from current measurements (and ratios) made with (perhaps) a thermocouple. *This is entirely satisfactory under special conditions where the impedance of the circuits are matched* (page 109); but this method cannot be used under other conditions without a correction, or errors will be introduced. To summarize, it is correct to determine *voltage* loss or

gain in decibels from voltage measurements; it is correct to determine *current* loss or gain in decibels from current measurements; but only when impedances of the circuits are matched is it correct to determine *power* loss or gain in decibels from voltage or current ratios without corrections (see footnote, page 95).

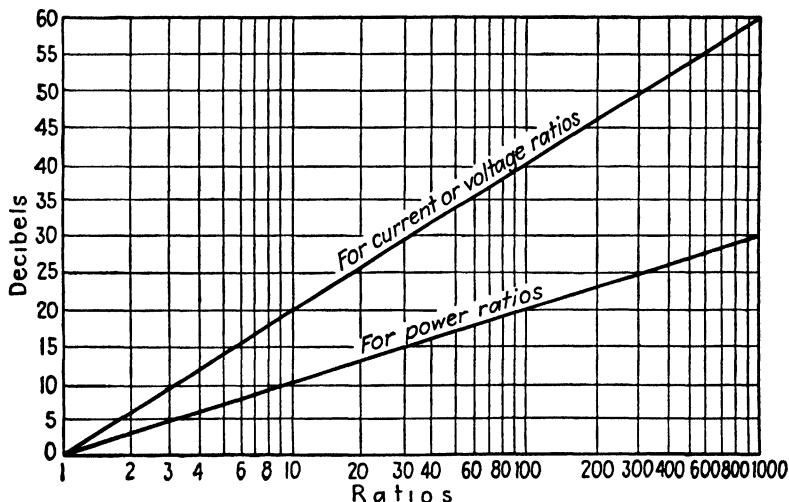



FIG. 56.—Curves for gain or loss in decibels from current, voltage, or power ratios.

Illustrative Problem.—The power entering a circuit under test is 1.0 milliwatt, and the power leaving the circuit is 0.05 milliwatt. Calculate the loss in decibels.

Solution.—Using Eq. (26), the power loss is

$$n = 10 \log_{10}\frac{P_1}{P_2} = 10 \log_{10}\frac{1.0}{0.05} = 10 \log_{10} 20 = 10 \times 1.301 = 13.01 \text{ decibels.}$$

These calculations can be checked on Fig. 56.

Power Levels.—The decibel is used also for expressing power levels, noise levels, transmission levels, volume levels, etc. By **power level** is meant the amount of power being transmitted past a point in a circuit. For such purposes it is necessary to establish an arbitrary reference, or **zero level.** (For example in measuring altitudes, sea level is used as the arbitrary reference, or zero level.) No arbitrary zero level has been accepted as standard. A power of 6.0 milliwatts has been used, and a power of 1.0 milliwatt is widely used now. The designation **dbm** sometimes is used to indicate that zero level is 1.0 milliwatt. In acoustics the arbitrary zero level is 10^{-16} watt.

Illustrative Problem.—The power input to a circuit is 0.01 milliwatt. What is the power level of the signal, using 1.0 milliwatt as zero level.

Solution.—Using Eq. (26), the power ratio is

$$n = 10 \log_{10}\frac{P_1}{P_2} = 10 \log_{10} \frac{1.0}{0.01} = 10 \log_{10} 100 = 10 \log_{10} 2 = 10 \times 2$$
$$= 20 \text{ decibels.}$$

Since the power input is less than the reference level, the power level is −20 decibels, that is, 20 decibels below the reference level of 1.0 milliwatt.

The Volume Unit.—If the magnitude of the audio-frequency signal impressed on the input of a radio transmitter is too great, the transmitting equipment may be overloaded and much distortion result. It is, accordingly, necessary to have a unit for measuring the magnitude or power level of the impressed speech or program signal. At one time the decibel was used for this purpose.

Difficulties were experienced with the early system because of confusion regarding the zero level to be used, and because the manufacturers used measuring instruments having different characteristics. As was shown in Chap. I, speech and music are quite complicated in nature, and unless the instruments used to measure the magnitude or volume of the speech power have the same electrical and mechanical characteristics, different instruments will indicate different values when they are connected in the *same* circuits. For instance, a highly damped instrument would be "sluggish" in operation, and a lightly damped instrument would tend to follow the instantaneous speech variations. Under such circumstances, confusion results.

About 1940, representatives of the organizations concerned standardized the characteristics for all radio volume-level measuring instruments. As a result, all such instruments now indicate substantially the same values in the same circuits. Thus the volume levels as measured by the telephone companies operating the broadcast network transmission lines agree with the volume level as measured at the radio stations receiving the programs.

The audio-frequency instrument used to measure radio program volume level is called a **volume-level indicator** (often abbreviated VI), and the unit of measure is the **volume unit** (often abbreviated VU). In defining zero volume level, the electrical and mechanical characteristics of the instrument and its method of use are specified, and the steady-state reference power is fixed at 1.0 milliwatt

flowing into a circuit of 600-ohms impedance.[1] The scales of the instruments are calibrated in volume units, these units being numerically equal to the number of decibels above zero volume level.

Maximum Power Transfer.—A radio system is composed of various units that pass the audio-frequency or radio-frequency signal along from one unit to another. Thus the speech signal from a remotely located microphone may be passed over telephone lines to the radio station, then through the audio-frequency speech-input equipment, then through the radio-frequency portion of the transmitting set, and then over a coaxial cable into the antenna, where the radio-frequency signal energy is broadcast into space. It is apparent that the successful operation of a radio system depends on the ability of one unit to draw signal power from the preceding unit and to pass this power on to the next. Such a process is called **power transfer,** and in communication **maximum power transfer** very often is important.

An illustration of maximum power transfer is given in Fig. 57. The power transferred from the primary cell, which has an open-circuit voltage E_{oc} and an internal resistance R_i, to the load resistor R_L is

$$P_L = I^2R_L = \left(\frac{E_{oc}}{R_i + R_L}\right)^2 R_L; \tag{30}$$

and the total power generated by the cell (part of which is lost in the internal resistance of the cell and the remainder is delivered to the load resistor) is

$$P_T = \left(\frac{E_{oc}}{R_i + R_L}\right)^2 (R_i + R_L) = \frac{E_{oc}^2}{R_i + R_L}. \tag{31}$$

The efficiency of the circuit in delivering or transferring power to the load resistor is

$$\text{Efficiency} = \frac{\text{power transferred}}{\text{power generated}} = \frac{\text{Eq. (30)}}{\text{Eq. (31)}} = \frac{R_L}{R_i + R_L}. \tag{32}$$

If various values of R_L (such as $R_L = 0.05$ ohm, $R_L = 0.1$ ohm, $R_L = 0.15$ ohm, etc.) are assumed and the power P_L transferred to the load and the efficiency for each condition are calculated, and

[1]Affel, H. A., H. A. Chinn, and R. M. Morris, A New "VI" and Reference Level, *Electronics*, Feb., 1939.

if the calculated values are plotted, curves such as Fig. 57*b* result. For the circuit under study, these curves show that *maximum power transfer occurs when the load resistance equals the internal resistance of the source, and that under these conditions the efficiency*

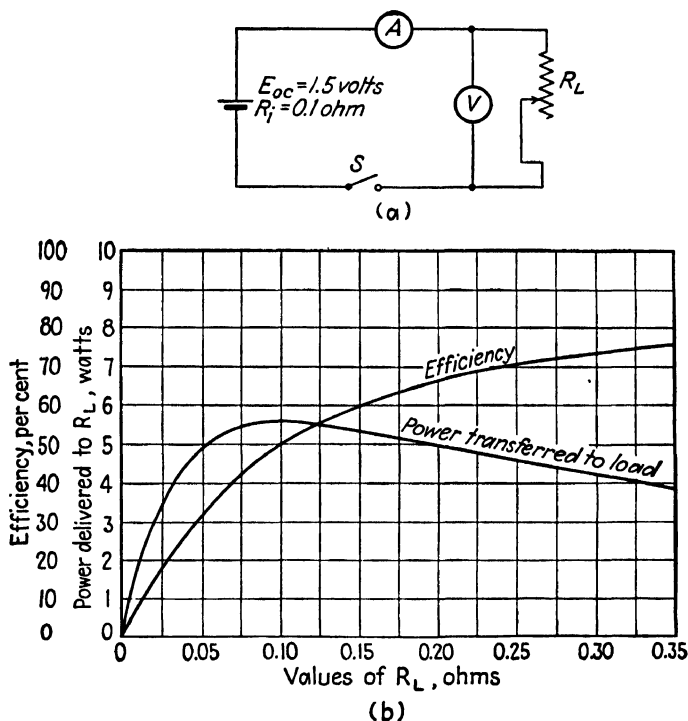



FIG. 57.—Circuit for studying power transfer, and curves showing that the maximum power transfer occurs when the load resistance R_L equals the internal resistance R_i. Under these conditions the efficiency is 50 per cent. If a generator of internal impedance Z_i is connected to a load Z_L, maximum alternating-current power will flow from the generator to the load when the resistance components of the two impedances are equal, and the reactive components are equal in magnitude and opposite in sign, that is, if one is inductive, the other is capacitive. Such impedances are known as conjugate impedances.

is 50 per cent. This figure of 50 per cent looks alarmingly low, and it must be remembered that in communication, factors other than power efficiency often are of the greatest importance.

The preceding illustration was for a direct-current circuit. In considering an alternating-current circuit, assume that a vacuum-tube oscillator having an open-circuit voltage of E_{oc} and an internal impedance Z_i is driving a load of impedance Z_L. If the internal

impedance of the oscillator is slightly inductive, then $Z_i = R_i + jX_i$. If the impedance of the load is $Z_L = R_L - jX_L$, and if $X_i = X_L$ in magnitude, the two reactances will be opposite, as indicated by the signs preceding the j terms, and will cancel; and the current will be controlled by the internal resistance of the oscillator and the resistance of the load. Under these conditions the circuit will follow the rules for maximum power transfer as previously explained for direct-current circuits.

Circuits having reactances that are equal in magnitude but opposite in sign are special cases. In general, a generator has an internal impedance of a certain resistance, reactance, and power-factor angle, and a load has a different resistance, reactance, and power-factor angle, and these cannot be changed at will.

As an example, the ordinary dynamic loudspeaker (page 35) used in radio-receiving sets is a low-impedance device, and it is driven by a vacuum-tube amplifier that is relatively a high-impedance device. Both the amplifier and loudspeaker are built in a prescribed way because they operate best when built that way, and their impedances cannot be changed at will. The question arises: What are the power-transfer characteristics under such conditions?

Both calculations and experimentation will show that when a high-impedance device drives a low-impedance device the power transfer is far from maximum. Also, calculations and experimentation will show that if some impedance-transforming circuit or device is inserted between the two units the power transfer is improved. To be specific, if an impedance-matching transformer (page 109) is connected between the amplifier and loudspeaker previously considered, then the input impedance to the loudspeaker (as measured through the connected transformer) will be altered, and with the correct transformer inserted, the loudspeaker can be made to draw much more power than it would draw without the transformer. For maximum power transfer under the conditions that the magnitudes but not the angles can be changed, the *magnitudes of the impedances of the two circuits should be made equal.*

Inductively Coupled Circuits.—As stated in the preceding section, a transformer can be connected between two devices to increase the power transfer. Such transformers are special types of inductively coupled circuits, which operate because of the mutual inductance considered on page 86. The properties of inductively coupled circuits will now be considered.

Two coils having mutual inductance are shown in Fig. 58. Assume for the moment that switch S is *open*. Then, in accordance with Eq. (25), page 86, the voltage induced in series with the turns of the secondary will be $E_S = 2\pi f M I_P$. The value of I_P will be determined by the impressed voltage and the resistance and inductance of the primary winding; provided that the various stray capacitances are negligible, *the open secondary has no effect on the primary.*

Now suppose that switch S is closed so that the induced secondary voltage E_S forces a current I_S through the secondary. Mutual inductance is a property that works both ways, and just as the magnetic effect of the current in the primary is felt in the secondary, so is the magnetic effect of the current in the secondary felt in the primary. The effect of the secondary current on the primary current is the same as if the primary circuit were altered by adding an impedance into the primary. This is called a **reflected impedance.** The impedance that the secondary reflects into the primary is given by

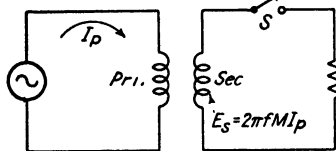



FIG. 58.—The current in the primary induces a voltage *in series* in the secondary winding. If the switch S is open, then the presence of the secondary has little or no effect on the primary. If the switch is closed, the secondary reflects an impedance into the primary, and this will affect the primary current.

$$Z_{\text{reflect}} = \frac{(2\pi f M)^2}{Z_S}. \tag{33}$$

The total primary impedance then becomes

$$Z'_P = Z_P + Z_{\text{reflect}}, \tag{34}$$

where Z_P is the primary impedance when the effect of the secondary is not considered.

In solving this equation for the reflected impedance, M is the mutual inductance in henrys, f is the frequency in cycles, and Z_S is the *entire* secondary impedance through which the secondary current flows. Furthermore, this impedance Z_S should be expressed in both magnitude and angle (or resistance and reactance, depending on the method used). This is because the phase angle of the secondary current will affect the nature of the reflected impedance. Thus in Eq. (33), if $Z_S = R_S + jX_S$, then $Z_{\text{reflect}} = R_{\text{reflect}} - jX_{\text{reflect}}$, because in dividing, the sign is changed (see

page 79). This means that an inductive reactance in the secondary appears as a capacitive reactance in the primary and a capacitive reactance in the secondary appears as an inductive reactance in the primary.

As written, Eq. (33) gives the *entire* impedance reflected into the primary by the effect of the secondary. From this the resistance and reactance components are obtained readily. It is, however, often of interest to know what these components will be prior to making a calculation for the *entire* impedance. For this purpose the following can be used:

$$Z_{\text{reflect}} = R_{\text{reflect}} - jX_{\text{reflect}} = \frac{(2\pi fM)^2 R_S}{R_S^2 + X_S^2} - j\frac{(2\pi fM)^2 X_S}{R_S^2 + X_S^2}. \quad (35)$$

In this equation f is the frequency in cycles per second, M is the mutual inductance in henrys, and R_S is the resistance and X_S is the reactance of the secondary circuit. This equation is obtained by multiplying both the numerator and denominator of Eq. (33) by $R_S - jX_S$, and by then separating the terms without and with letter j preceding. An example of the use of these equations will be given now.

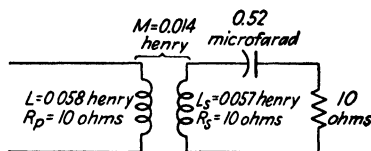



FIG. 59.—Circuit for studying the impedance that the secondary reflects into the primary.

Illustrative Problem.—Calculate the impedance reflected into the primary of the circuit of Fig. 59 at a frequency of 925 cycles, and calculate the final primary impedance.

Solution.—Step 1. Calculate the total series impedance of the secondary, using Eq. (21), page 76.

$$Z_S = 20 + j\left(6.28 \times 925 \times 0.057 - \frac{1}{6.28 \times 925 \times 0.52 \times 10^{-6}}\right)$$
$$= 20 \text{ ohms (approximately).}$$

Step 2. Calculate the impedance reflected into the primary, using Eq. (33).

$$Z_{\text{reflect}} = \frac{(6.28 \times 925 \times 0.014)^2}{20} = \frac{(81.2)^2}{20} = \frac{6600}{20} = 330 \text{ ohms.}$$

Step 3. The primary impedance under these conditions can be found from Eq. (34).

$$Z_P' = (10 + 330) + j(6.28 \times 925 \times 0.058) = 340 + j337 \text{ ohms.}$$

It is important to note that in the preceding example the 10 ohms of resistance connected to the secondary of the circuit causes a total of 340 ohms of resistance to appear in the primary. This

is an example of an impedance change or impedance transformation.

Coefficient of Coupling.—Two circuits may be so situated that the magnetic flux produced by the current in the primary links but *little* with the turns of the secondary. Such circuits are **loosely coupled circuits.** On the other hand, two circuits may be so situated that *much* of the flux caused by primary links with the secondary. Such circuits are **closely coupled circuits.** The **coefficient of coupling** is given by the relation

$$k = \frac{M}{\sqrt{L_P L_S}}, \tag{36}$$

where k is less than unity (and may be expressed as a per cent by multiplying by 100), when L_P is the primary self inductance, L_S is the secondary self-inductance, and M is the mutual inductance between them. These usually are expressed in henrys.

An audio-frequency transformer, or a 60-cycle power transformer, are examples of closely coupled circuits in which the coupling approaches 100 per cent. On the other hand, air inductors (coils), such as the ones which are used extensively in radio, are often loosely coupled. Coils quite far apart are loosely coupled; also, coils at right angles are loosely coupled.

A variable inductor often is needed for audio-frequency and radio-frequency circuits. As mentioned on page 65, such an inductor can be realized by varying the number of turns (changing the taps) on a coil. An inductor of this kind is not continuously variable, however, and also as the number of turns is changed the resistance is varied; often this is an objectionable feature.

A variable inductor called an **inductometer**, or a **variometer**, can be made of two coils, one fixed in position and the other arranged for rotation. If the two coils are connected in series so that the coils *aid*, then the inductance between the series terminals is the sum of the primary, secondary, and mutual inductive effects. The term "aiding" means this: The mutual inductive effects cause the current in the stationary winding to induce a voltage in the movable winding; also, the same series current in the movable winding induces a voltage in the stationary winding. The winding can be connected *aiding* so that these effects add, or *opposing* so that these effects cancel each other.

If the windings are connected "aiding," then the self and mutual

effects add, and a variable inductor is available, because one winding can be moved with respect to the other, the coefficient of coupling can be varied at will, and from Eq. (35)

$$M = k\sqrt{L_P L_S}. \tag{37}$$

Of course the inductance of this device cannot be reduced to zero because the primary and secondary inductance always are in series. The inductance can be varied over a wide range, however, and for all settings *the series resistance remains substantially the same* for a well-constructed device. This often is a very desirable feature.

Inductively Coupled Resonant Circuits.—These circuits often are used in radio, an example being given in Fig. 60. The magnitude of the voltage maintained across points 1–2 is constant and the frequency of the voltage is variable. Also, both circuits contain the same capacitance C_P and C_S, the same inductance L_P and L_S, and the same resistance R_P and R_S. These two circuits are, therefore individually tuned to the same resonant frequency $f_r = 1/(2\pi\sqrt{L_P C_P})$, or $f_r = 1/(2\pi\sqrt{L_S C_S})$. To analyze the circuit of Fig. 60, the input impedance at the driving points 1–2 will be considered at three different frequencies.

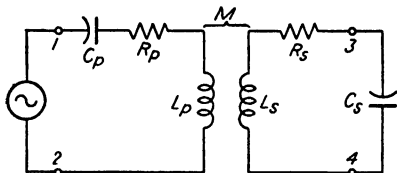



FIG. 60.—Inductively coupled resonant circuits.

Below frequency f_r the capacitive reactances will exceed the inductive reactances in each of the individual circuits. But a capacitive reactance in the secondary will be coupled into the primary as an inductive reactance (page 103), and this coupled reactance will add to that already present. At some frequency below f_r, the amount of inductive reactance coupled into the primary will be just sufficient to make the total primary inductive reactance equal to the primary capacitive reactance. At this frequency (below f_r) resonance will occur between the driving points 1–2, and the input current will be in phase with the voltage across points 1–2.

Above frequency f_r the inductive reactances will exceed the capacitive reactances in each of the individual circuits. But an inductive reactance in the secondary will be coupled into the primary as a capacitive reactance, and this coupled reactance will add to that already present. At some frequency above f_r, the

amount of capacitive reactance coupled into the primary will be just sufficient to make the total primary capacitive reactance equal to the primary inductive reactance. At this frequency (above f_r) resonance again will occur between the driving points 1–2, and the input current again will be in phase with the voltage across points 1–2.

A tuned radio-frequency transformer, such as used in a radio-receiving set. This is an example of inductively coupled resonant circuits with loose coupling. The primary and secondary windings are shown within the cut-away metal shield. The two variable condensers for tuning the primary and secondary are at the top. The form on which the coils are placed is nonmagnetic. (*Meissner Manufacturing Division, Maguire Industries, Inc.*)

At frequency f_r the inductive reactances and the capacitive reactances are equal in each of the individual circuits because they are tuned to resonance at this frequency. At the frequency f_r the equivalent impedance of the secondary is resistance only, and this couples resistance only into the primary. Since the primary is in resonance, its equivalent impedance is resistance only. The resistance coupled into the primary by the secondary increases the total primary resistance, but it does not affect the value of the resonant frequency f_r.

The effect of the coupling on the phenomena just discussed is of importance. Theoretically, there are always *three* points at which the driving voltage (across points 1–2 of Fig. 60) and the input current are in phase. But if the coefficient of coupling and the mutual inductance are reduced from large values to small values, these three points move together (that is, occur at frequencies more nearly the same), and as the coefficient of coupling approaches a critical value, the three points become indistinguishable.

When circuits similar to Fig. 60 are used in radio, the current through the secondary and the voltage across the secondary at points 3–4 often are of interest. The voltage between points 3–4 is equal to the $I_S X_C$ drop across condenser C_S. On the assumption that X_C is approximately the same, as it will be for frequencies close to $f_r = 1/(2\pi\sqrt{L_S C_S})$, the voltage between points 3–4 will depend on the secondary current. Also on the assumption that

the Q of the coils remains substantially constant, when the frequency is varied, curves for secondary current, such as the ones shown in Fig. 61, result.

For low values of k (and resulting low mutual inductance) the voltage induced in the secondary [Eq. (25), page 86], the secondary current, and the voltage drop $I_S X_C$ between points 3–4 are low; also, the resonance points coincide and the curve is steep. Such a circuit is *loosely coupled* and *sharply tuned.*

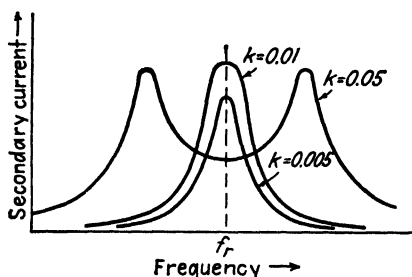



FIG. 61.—For various coefficients of coupling, the relative magnitudes of the secondary currents of Fig. 60 will be as shown. The frequency of resonance is f_r.

For some value of k called the **critical coupling,** at the resonant frequency f_r the resistance [Eq. (35)] coupled into the primary will equal R_P. This is the condition for maximum power transfer to the secondary (page 101), because the equivalent circuit of Fig. 60 now becomes a generator connected to R_P and R_{reflect} in series, and any power delivered to R_{reflect} is in reality delivered to the secondary. This critical coupling is [from Eq. (33)] when $(\omega M)^2/R_S = R_P$, or when

$$\omega M = \sqrt{R_P R_S}. \tag{38}$$

From Eq. (11), page 49, $Q_P = \omega L_P/R_P$, and $Q_S = \omega L_S/R_S$; also, from Eq. (36), $k = M/\sqrt{L_P L_S}$. Making these substitutions in Eq. (38) gives

$$\omega k \sqrt{L_P L_S} = \sqrt{R_P R_S}, \qquad \text{and} \qquad k \doteq \sqrt{\frac{R_P}{\omega L_P} \frac{R_S}{\omega L_S}} \qquad \text{or}$$

$$k = \frac{1}{\sqrt{Q_P Q_S}}. \tag{39}$$

From this equation it is seen that if each of two circuits has a $Q = 100$, then the critical coupling will be $k = 0.01$, or 1.0 per cent. As Fig. 61 indicates, at this critical coupling the current

and voltage curves will be well rounded off, that is, not especially sharp as with lower values of k.

For high values of k, above the critical coupling, the three resonance points show up. The two peaks, off the resonance frequency $f_r = 1/(2\pi\sqrt{LC})$, are higher than at the value f_r. The reason for this is that, although resonance is produced, the value of the resistance component coupled into the primary is not so great as at the frequency f_r. Hence, a larger primary current flows (than at the frequency f_r), and the secondary voltage and current are greater than at f_r.

The Closely Coupled Transformer.—The two coils just considered may be classed as a loosely coupled transformer. Because of the loose coupling between the primary and secondary, it is necessary to resort to mutual inductance and reflected impedance theory in studying the circuit.

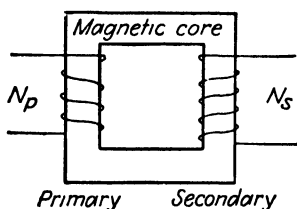



FIG. 62.—A closely coupled transformer. Because of the excellent path provided by the magnetic core, it may be assumed that all the magnetic lines of force produced by the primary link with the secondary.

For 60-cycle and other power frequencies, and for audio frequencies, closely coupled circuits (or transformers) are used. The primary and secondary windings are placed on a common core of some magnetic material such as silicon-steel laminations or a core of powdered and compressed Permalloy, an alloy often of nickel and iron.

In these transformers, shown schematically in Fig. 62, the cores are so good and the windings are so well arranged that as an approximation all the magnetic flux produced by the primary current is assumed to pass through and to link with the secondary. Other justifiable assumptions are that the primary and secondary windings are lossless (resistance negligible) and that there are no hysteresis and eddy-current losses in the core. If these assumptions are made, then the power $E_P I_P \cos \theta_P$ put into the primary must equal the power $E_S I_S \cos \theta_S$ that flows out of the secondary. Such an ideal transformer would not change the phase angle, and thus $E_P I_P = E_S I_S$.

When such a transformer, with no primary or secondary resistance, is connected to a source of voltage, the back voltage in

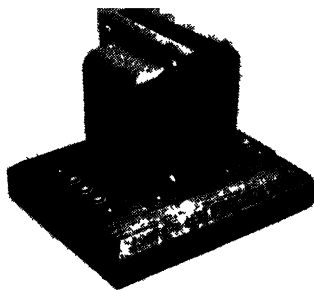
duced in the primary must equal the impressed voltage. This back voltage is induced by the magnetic flux in the core, and this same flux links both the primary and secondary windings. Since the magnitude of an induced voltage is directly proportional to both the magnetic flux and the number of turns, it follows that the primary and secondary voltages are directly proportional to the primary and secondary turns. Thus

$$\frac{E_S}{E_P} = \frac{N_S}{N_P}, \quad \text{or} \quad E_S = \frac{N_S E_P}{N_P}, \quad \text{or} \quad E_P = \frac{N_P E_S}{N_S}. \quad (40)$$

Furthermore, since, as previously stated, the volt-ampere product in the primary $E_P I_P$ must equal the volt-ampere product in the secondary $E_S I_S$, if the voltage is increased by a ratio N_P/N_S, the current must be decreased by the same amount. Therefore

$$\frac{I_S}{I_P} = \frac{N_P}{N_S}, \quad \text{or} \quad I_S = \frac{N_P I_P}{N_S},$$

$$\text{or} \quad I_P = \frac{N_S I_S}{N_P}. \quad (41)$$

An example of a closely coupled transformer is the audio-frequency transformer shown This is the type sometimes used in vacuum-tube amplifiers The transformer is mounted on a wooden block for laboratory testing purposes.

From these last two equations it follows that if a transformer increases voltage it decreases current, and if it decreases voltage it increases current. In 60-cycle power studies the transformer is regarded as a voltage and current changer. Although this same viewpoint is used in radio also, particularly in power-supply design, the transformer often is regarded as an **impedance changer** or an as **impedance matcher.** This is especially true in audio amplifier design where, for example, the power-output transformer is used to match an 8-ohm loudspeaker voice coil to the high-resistance plate circuit of a vacuum tube. The reason for this impedance changing and matching viewpoint will be given now.

The **driving-point impedance,** or impedance "looking into" the transformer primary, equals the alternating voltage connected *across* the primary terminals divided by the alternating current

flowing into the terminals. That is, $Z_P = E_P/I_P$. Substituting the appropriate values from Eqs. (40) and (41) gives

$$Z_P = \frac{(N_P E_S/N_S)}{(N_S I_S/N_P)} = \frac{N_P E_S}{N_S} \times \frac{N_P}{N_S I_P}$$

$$= \left(\frac{N_P^2}{N_S}\right)\frac{E_S}{I_S} = \left(\frac{N_P}{N_S}\right)^2 Z_S. \qquad (42)$$

This equation states that if an impedance of Z_S is connected to the secondary of a closely coupled lossless transformer, then the impedance measured between the primary terminals will be equal to the square of the turns ratio times the secondary impedance. Thus if there are more primary than secondary turns, the impedance will be *increased*, and if the reverse is true the impedance will be *decreased*. From Eq. (42)

$$Z_P = \left(\frac{N_P}{N_S}\right)^2 Z_S, \qquad \text{and} \qquad \frac{N_P}{N_S} = \sqrt{\frac{Z_P}{Z_S}}. \qquad (43)$$

It is again pointed out that a transformer, if it is well built, causes negligible change in the phase angle between the voltages and currents in their respective windings. (If this were not true, then 60-cycle power could be drawn from the supply mains and a transformer could be arranged so that on the primary side the current and voltage would be out of phase, and the watthour meter would not properly register the energy taken.)

In considering maximum power transfer on page 101, it was stated that for maximum power transfer between two circuits in which the magnitudes of the impedances but not the angles could be changed, maximum power transfer would result when the magnitudes of the two impedances were equal.

Illustrative Problem.—A dynamic loudspeaker having an 8-ohm voice coil is to be connected to an amplifier designed to work into a 10,000-ohm load. How should the two be connected?

Solution.—From Eq. (43), $N_P/N_S = \sqrt{\frac{10{,}000}{8}} = \sqrt{1250} = 35$ (approximately).

Thus a power-output transformer having a turns ratio of about 35 primary turns to 1 secondary turn should be used. A word of warning should be given: Any transformer having this turns ratio will not do. The transformer must be designed to work between such impedances, must have the required frequency response, and must have adequate power-handling capacity.

Equivalent Series and Parallel Circuits.—In many instances in radio it is desired to determine the series circuit that will be equivalent to a given parallel circuit. Sometimes the reverse process is to be performed; that is, it is desired to determine the parallel circuit which is equivalent to a given series circuit. Simple two-terminal circuits (such as a resistor and a reactor in series, or in parallel) are equivalent if they have identical input or driving-point impedances and therefore draw the same current at the same phase angle when the same voltage source is connected across them. Ordinarily, they are equivalent at *only one frequency or at one narrow band of frequencies.*

FIG. 63.—Circuits for determining series and parallel equivalent circuits.

To derive the equations for determining the parallel circuit that is equivalent to a given series circuit (or vice versa) the currents drawn by the two circuits are equated.[1] Thus consider the two circuits of Fig. 63. One circuit is equivalent to the other if the currents drawn are identical when the same voltages are applied. The current taken by the series circuit is $I_S = E/(R_S + jX_S)$. The current taken by the parallel circuit is $I_P = E/R_P + E/(jX_P)$. Since these two currents must be identical for the circuits to be equivalent,

$$\frac{E}{R_S + jX_S} = \frac{E}{R_P} + \frac{E}{jX_P} \quad \text{and} \quad \frac{1}{R_S + jX_S} = \frac{1}{R_P} + \frac{1}{jX_P}. \tag{44}$$

This equation is solved by algebraic means to eliminate the denominators. Next, the real, or in-phase, terms are equated separately, and the reactive, or out-of-phase, terms are equated separately. This is done because the respective components of the two currents must be identical for the circuits to be equivalent. The next step is to make substitutions so that the expression for parallel circuit terms will be in series circuit terms, and vice versa.

For changing from a series to a parallel circuit the equations are

$$R_P = R_S\,(1 + Q^2) \qquad \text{and} \qquad X_P = \frac{R_P}{Q} \tag{45}$$

where R_P and X_P are the desired units of a *parallel* circuit that is

[1] This discussion is based on an excellent article by W. N. Tuttle, *General Radio Experimenter*, Vol. 20, No. 8, January, 1946.

equivalent to a series circuit composed of R_S and X_S. The value of Q in Eq. (45) is $Q = X_S/R_S$.

For changing from a parallel to a series circuit the equations are

$$R_S = \frac{R_P}{1 + Q^2} \quad \text{and} \quad X_S = R_S Q, \tag{46}$$

where R_S and X_S are the desired units of a *series* circuit that is equivalent to a parallel circuit composed of R_P and X_P. The value of Q in Eq. (46) is $Q = R_P/X_P$.

In using the equations just given, it is important to note that the Q values are *computed in different ways*, depending on whether data for series circuits or for parallel circuits are given. It is possible to write the equations so that the dissipation factor D instead of the energy storage factor Q is used (page 49). The resistance and reactance values in these equations are in ohms.

Illustrative Problem.—A circuit is composed of a 10-ohm resistor and a coil having 10 ohms reactance in series. The coil is assumed to be lossless. Find the equivalent parallel circuit.

Solution.—$Q = X_S/R_S = 10/10 = 1$.

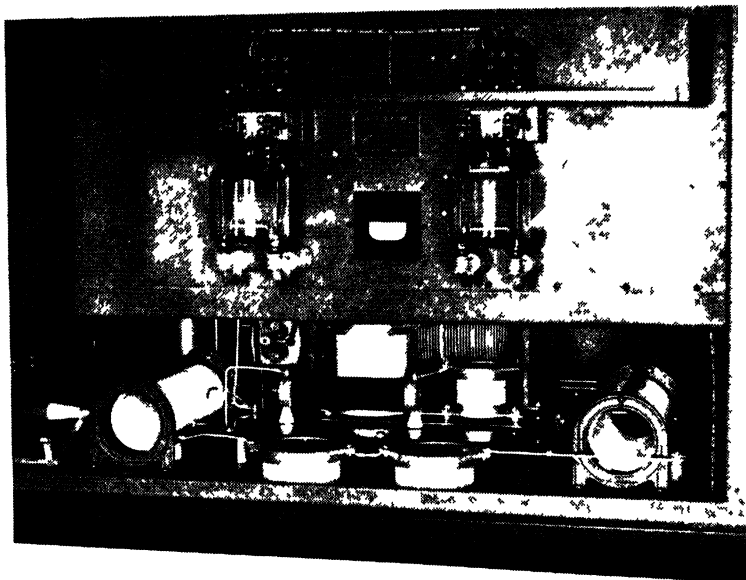
From Eq. (45)

$$R_P = R_S(1 + Q^2) = 10(1 + 1^2) = 20 \text{ ohms} \quad \text{and}$$
$$X_P = R_P Q = 20 \times 1 = 20 \text{ ohms}.$$

Thus a circuit of 20 ohms resistance in *parallel* with 20 ohms inductive reactance is equivalent to a circuit of 10 ohms resistance in *series* with 10 ohms inductive reactance.

Impedance-transforming Circuits.—At several points in this chapter it has been explained that circuits must be matched in certain ways to obtain maximum power transfer from one device to another, or from one circuit portion to another. It has also been explained that in the practical problem the impedances of apparatus seldom can be changed at will to obtain maximum power transfer. There are many ways, however, in which the impedance of a device or circuit can be *transformed* so that it will properly match a load and draw maximum power from the load.

Impedances can be transformed and matched to other circuits by connecting impedance-transforming circuits, or networks, between the two circuits to be matched. The closely coupled transformer described in this chapter is one method widely used. As was brought out, however, this transformer cannot change the phase angle between the voltage and current, and sometimes it is desired to use an impedance-matching network that will change the phase

angle and perhaps eliminate it entirely. Also, the loosely coupled circuit was shown in the example on page 102 to have impedance-transforming properties. Two useful methods of impedance transformation will be considered now. For simplicity, the impedances to be transformed will be *pure resistances*. If they are not pure resistances but contain some reactance, sufficient reactance of

An impedance-transforming circuit of a radio transmitter is shown above the two tubes. This matches the output circuit of the final power-amplifying tubes to a coaxial cable, so that maximum radio-frequency power flows from the transmitter to the cable and then to the transmitting antenna. (*Collins Radio Co.*)

the opposite sign can be connected in series with them so that they are then effectively pure resistances.

To increase a value of resistance the circuit of Fig. 64*a* can be used. The resistance R can be a resistor actually inserted in series with the coil, or it can be a value reflected in series with the coil in accordance with coupled-circuit theory. Of course R might also be the resistance of the coil.

The branch of Fig. 64*a* composed of L and R ***in series*** can be changed to an equivalent parallel circuit by using Eq. (45). The circuit then becomes L' and R' of Fig. 64*b*, where

$$R' = R(1 + Q^2) \quad \text{and} \quad X'_L = 2\pi f L' = \frac{R'}{Q}, \tag{47}$$

and $Q = X/R = 2\pi fL/R$. Then, if a value of C is selected and adjusted so that it takes as much leading current as L' takes lagging current, resonance is obtained, and the impedance Z between the input terminals is a *high* value of pure resistance R', as given by Eq. (47). From this equation it is apparent that the value of R' can be varied by selecting the proper value of L. Thus a resistance of value R can be *increased* to a higher value R'. In Eq. (47) the values should be in ohms, henrys, farads, and cycles per second.

Illustrative Problem.—The input impedance to a device is 100 ohms pure resistance. For impedance-matching reasons this should be increased to 500 ohms pure resistance. The frequency is 5,000,000 cycles.

Solution.—Step 1. The circuit of Fig. 64*a* will be used for this purpose. Using Eq. (47), $R' = R(1 + Q^2)$, where $Q = 2\pi fL/R$. Thus, $R' = R[1 + (2\pi fL/R)^2]$, or $500 = 100[1 + (6.28 \times 5 \times 10^6 L/100)^2]$. Solving for L gives $L = 6.3 \times 10^{-6}$ henry (approximately). This is the value of the coil that should be connected *in series* with the 100 ohms as indicated in Fig. 64*a*.

Step 2. Calculate the equivalent inductive reactance of coil L when transformed into the equivalent circuit of Fig. 64*b*. Equation (47) also is used for this. $X_L = 2\pi fL' = R'/Q = R'/\left(\frac{2\pi fL}{R}\right)$

$$= 500/\left(\frac{6.28 \times 5 \times 10^6 \times 6.3 \times 10^{-6}}{100}\right) = 253 \text{ ohms.}$$

Step 3. Calculate the value that C of Fig. 64 must have. The capacitive reactance of C must equal the inductive reactance of L', or 253 ohms. Thus $1/(2\pi fC) = 253$, and $C = 1/(2\pi f \times 253)$

$$= 1/(6.28 \times 5 \times 10^6 \times 253) = 125 \text{ micromicrofarads.}$$

This solution can be checked by using Eq. (23), page 82, to see if a branch composed of a 6.3-microhenry coil and a 100-ohm resistor, when in parallel with a 125-micromicrofarad condenser, gives an equivalent input impedance of 500 ohms at 5,000,000 cycles. Such a check shows that these values are substantially correct. In a circuit such as Fig. 64*a* the final adjustment is made by varying the condenser.

To decrease a value of resistance the circuit of Fig. 65*a* can be used. The resistance R to be transformed is connected in parallel with a coil L. This parallel portion can be changed to an equivalent series circuit by Eq. (46). The circuit then becomes that of Fig. 65*b*, where

$$R' = \frac{R}{1 + Q^2} \quad \text{and} \quad X'_L = 2\pi fL' = R'Q, \tag{48}$$

and $Q = R/X = R/2\pi fL$. Then if a value of C is selected and adjusted so that it neutralizes the inductive reactance of L', the impedance Z between the input terminals is a *low* value of pure resistance, as given by Eq. (48). Thus a value of resistance R can be *decreased* to a lower value R'. In Eq. (48) the values should be in ohms, henrys, farads, and cycles per second.

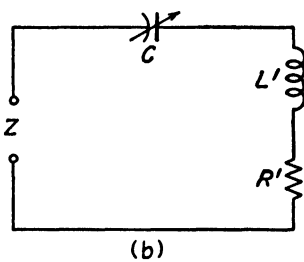
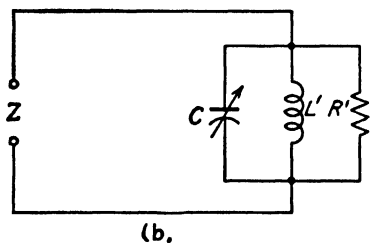
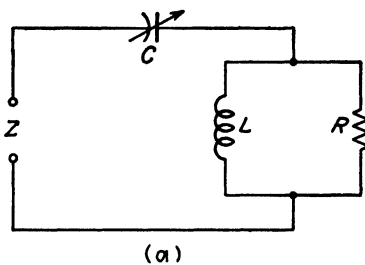
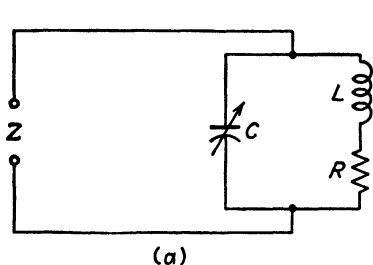



FIG. 64.—An impedance-transforming circuit (a), and the equivalent circuit (b). The circuit (a) is used for increasing the value of the resistance R.

FIG. 65.—An impedance-transforming circuit (a), and the equivalent circuit (b). The circuit (a) is used for decreasing the value of the resistance R.

Illustrative Problem.—The input impedance to a device is 500 ohms pure resistance. For impedance matching reasons this should be decreased to 100 ohms pure resistance. The frequency is 10,000,000 cycles.

Solution.—Step 1. The circuit of Fig. 65*a* will be used for this purpose. Using Eq. (48), $R' = R/(1 + Q^2)$, where $Q = R/2\pi fL$. Thus, $R' = \dfrac{R}{1 + \left(\dfrac{R}{2\pi fL}\right)^2}$ or $100 = \dfrac{500}{1 + \left(\dfrac{500}{6.28 \times 10^7 \times L}\right)^2}$. Solving for L gives $L = 3.95 \times 10^{-6}$ henry. This is the value of the coil that should be connected *in parallel* with the 500 ohms as in Fig. 65*a*.

Step 2. Calculate the equivalent inductive reactance of coil L when transformed into the equivalent circuit of Fig. 65*b*. Equation (48) also is used for this. $X_L' = 2\pi fL' = R'Q = R'R/2\pi fL = 100\ (500/6.28 \times 10^7 \times 3.95 \times 10^{-6}) = 201$ ohms.

Step 3. Calculate the value that C of Fig. 65 must have. The capacitive

reactance must equal the inductive reactance of L', or 201 ohms. Thus

$$\frac{1}{2\pi fC} = 201 \quad \text{and} \quad C = \frac{1}{2\pi f \times 201} = \frac{1}{6.28 \times 10^7 \times 201}$$
$$= 79.3 \text{ micromicrofarads.}$$

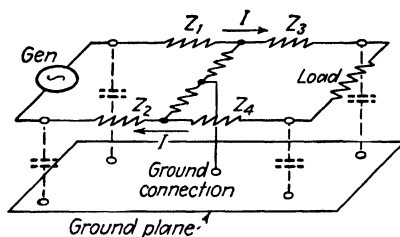
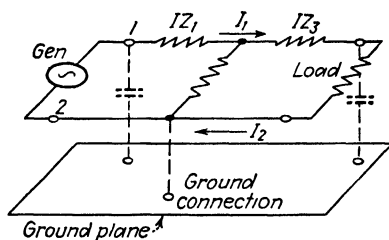
This solution can be checked by using the method considered on page 85 for series-parallel circuits. Such a calculation shows that when a 500-ohm resistor and a 3.95-microhenry coil are connected in parallel, and the combination is placed in series with a 79.3-micromicrofarad condenser, that this series-parallel combination (Fig. 65*a*) does have an input impedance of substantially 100 ohms as desired.

(a)-A balanced network having equal impedances Z in each line wire

(b)-An unbalanced network having unequal impedances in each line wire

FIG. 66.—Circuits which are balanced and unbalanced with respect to ground.

Balanced and Unbalanced Circuits.—The circuits and networks used in radio are of two general types: those which are **balanced** with respect to ground, and those which are **unbalanced** with respect to ground. The term "ground" may be the earth itself, or some object such as a metal table top, or electric conduit, or the water system.

A balanced circuit is of the type shown in Fig. 66*a*. If the impedances in each side of the circuit are the same, and if the stray capacitances to ground are the same, then the current in each line wire between the generator and the load will be identical, and the

circuit is said to be balanced. If the circuit is to be grounded, the ground should be placed as indicated at the center point of a shunt impedance.

An unbalanced circuit is of the type shown in Fig. 66*b*. In an unbalanced circuit the impedance of one side is different from that of the other. Although the stray capacitances to ground still may be about the same, the series IZ drops in the two sides will be different, and different currents will flow through the stray capacitances to ground. Because of this, the currents in the two line wires will be different, and the circuit is said to be unbalanced.

In radio, both balanced and unbalanced circuits are used. Push-pull vacuum-tube amplifiers (page 333) are balanced, but single-tube amplifiers (page 324) are unbalanced. Two-wire transmission lines from a transmitter to the antenna are balanced, but coaxial cables are unbalanced.

This is a very important principle to remember: Balanced circuits and unbalanced circuits are distinctly of different types. *When a balanced circuit is to be connected to an unbalanced circuit, the connection should be made through a transformer* (that is, inductively through an audio-frequency transformer in speech-input equipment, or through two inductively coupled coils for radio-frequency apparatus). Unbalanced measuring and test equipment usually will operate incorrectly when used with balanced circuits, and vice versa.

Shielding.—Circuit elements, such as resistors, inductors, and capacitors, usually are carefully shielded in radio test and other equipment so that one element will not be coupled electrically with another. Sometimes the entire circuit is shielded. In radio receivers, the transformers are shielded by placing them in metal cans. There are many applications of shielding.

At low frequencies, such as at 60 cycles and the lower audio frequencies, coils must be shielded by placing them in containers made of good magnetic material, such as soft iron. Then, the container *conducts* stray magnetic lines of force *around* the coil within, and in this way keeps the stray field from affecting the coil and inducing a stray voltage in it. Also, the shield of magnetic material keeps the magnetic lines of force produced by the coil from leaving the vicinity of the coil and affecting other circuits.

At the higher audio frequencies, and at radio frequencies, a container of some good conductor, such as copper or aluminum,

prevents stray magnetic fields from inducing a stray voltage in a coil, and also keeps the coil from affecting other circuits. The action is different from that of the preceding paragraph, because copper and aluminum are essentially nonmagnetic. The shielding action is caused by the eddy currents that the rapidly reversing fields induce in the conducting container.

The rapidly changing stray magnetic fields cause eddy currents to flow in the container, and these produce magnetic fields that neutralize the stray fields and prevent them from penetrating the container. By a similar action, the magnetic field of a coil within the container is kept from leaving the container. For very high radio frequencies these eddy currents flow very near the container surface, and the outer surface layer shields the coil from external magnetic fields and the inner surface layer shields external circuits from magnetic fields produced by the coil.

Metal containers also shield circuits from stray electric fields, because the electric lines of force terminate on charges at the metal surface, and the lines do not penetrate within. Similarly, a metal container surrounding a device will shield other circuits from the electric field produced by that device. Whether or not a shield should be grounded depends on circumstances. In experimental work it may be well always to ground a circuit first through a voltmeter or a fuse to ascertain if a voltage exists between the shield and ground.

In general, *unbalanced circuits must be carefully shielded.* As an illustration, the outside sheath of a coaxial cable provides the necessary shielding. On the other hand, *balanced circuits often are operated without shielding.* For example, a two-wire transmission line often is unshielded. The reason for this is as follows: For a balanced two-wire circuit stray pickup *is* experienced; but the stray voltages induced in each line wire are essentially equal, are in the same direction, and largely cancel. To equalize further these induced voltages, the wires of a transmission line often are **transposed,** or interchanged in position, at regular intervals. Two-wire (insulated) circuits often are twisted together to interchange the positions of the two wires and further equalize the induced voltages. This also ensures good balance to ground and reduces interference. The twisted wires sometimes are enclosed in a copper braid that shields the conductors within. At the high frequencies

used in radio, interference is particularly bothersome, and it may be necessary to make use of all available means of reducing interference.

SUMMARY

Radio transmitters and receivers are composed of various sections that accomplish certain operations. The electric signals to be transmitted or received are passed from one section to another, and the success of this depends on the impedance relations between the sections. Impedance transformations often are necessary to match the sections.

The decibel is used to express the increase or decrease in signal strength when a radio signal passes through circuits. For measurements of *power* ratios the equation is

$$n = 10 \log_{10} \frac{P_1}{P_2},$$

where P_1 is the power input and P_2 the power output for measuring a signal decrease or loss in decibels. For power increases or gain P_1 is the power output and P_2 the power input.

The decibel also is used to express the increase or decrease of the voltage of radio signals. For the measurement of *voltage* ratios the equation is

$$n = 20 \log_{10} \frac{E_1}{E_2},$$

where E_1 is the voltage input and E_2 is the voltage output for measuring a signal decrease or loss in decibels, and E_1 is the voltage output and E_2 the voltage input for measuring a voltage increase or gain. Current decreases and increases can be expressed in decibels by using the same equation.

The power level, or the amount of power being transmitted past a point in a circuit, is measured in decibels above an arbitrarily selected reference or zero level. A zero level of 1.0 milliwatt is common, but other levels have been used.

The volume unit is used in radio to express the strength of the audio-frequency signal. Special instruments are used to indicate the volume level.

Maximum power transfer is obtained between circuits of resistance only, when the internal resistance of the source equals the resistance of the load. For circuits containing resistances and reactances, maximum power transfer results when the resistances are equal and the reactances are equal and opposite, that is, one inductive and one capacitive. If the magnitudes but not the phase angles of two impedances can be altered, maximum power transfer occurs when the magnitudes of the two circuits are equal.

When two circuits are inductively coupled, the effect of a closed secondary is to reflect an impedance into the primary of the amount

$$Z_{\text{reflect}} = \frac{(2\pi fM)^2}{Z_S}.$$

This reflected impedance adds to the primary impedance to give the total effective impedance.

If most of the magnetic flux produced by the primary links with the second-

ary, two coils are closely coupled. If the reverse is true, they are loosely coupled. The coefficient of coupling is

$$k = \frac{M}{\sqrt{L_P L_S}}.$$

Inductively coupled circuits often are used primarily as impedance transformers.

The transformer used in 60-cycle power-supply systems and in audio-frequency circuits is an example of circuits that have a coefficient of coupling approaching unity, or 100 per cent. These transformers are used extensively for impedance transforming and matching.

A parallel circuit exists that will draw the same current at the same voltage and is hence equivalent to a series circuit; also, the reverse of this statement is true. The equations for the parallel circuit in terms of the series elements are

$$R_P = R_S\,(1 + Q^2) \qquad \text{and} \qquad X_P = \frac{R_P}{Q}, \qquad \text{where} \qquad Q = \frac{X_S}{R_S}.$$

For changing from a parallel to a series circuit the equations are

$$R_S = \frac{R_P}{1 + Q^2} \qquad \text{and} \qquad X_S = R_S Q, \qquad \text{where} \qquad Q = \frac{R_P}{X_P}.$$

Coupled circuits have impedance-transforming properties. Such changes also can be accomplished with simple parallel and series-parallel circuits.

Circuits are balanced or unbalanced with respect to ground. Such circuits should be interconnected only through inductively coupled circuits, such as transformers.

At the high frequencies used in radio, a thin container of some good conducting material, such as copper or aluminum, shields from both magnetic and electric fields.

QUESTIONS

1. In what ways is the decibel used?
2. Under what conditions may voltage measurements be used to measure power losses in decibels?
3. What is meant by power level? How is the decibel used to measure power level?
4. Why must a special instrument be used to measure volume level in speech-input equipment?
5. Why is maximum power transfer of importance?
6. When does maximum power transfer occur in direct-current circuits? In alternating-current circuits?
7. When are two circuits inductively coupled?
8. What is meant by reflected impedance? How is it computed?
9. What is the relation between the sign of the reactance in the secondary and the sign of the reactance in the reflected impedance? Use Eq. (33) to explain your answer.
10. What is an important advantage of the variometer principle of obtaining a variable inductance?
11. What determines when two circuits are critically coupled?

12. Why is the transformer used so extensively in audio circuits?

13. In selecting a transformer for audio-frequency circuits, what factors other than turns ratio are very important?

14. Why is it true that for every series circuit an equivalent parallel circuit exists?

15. What is meant by an impedance-transforming circuit? Why are thev used?

16. What is the difference between a balanced and unbalanced circuit?

17. How should balanced and unbalanced circuits be interconnected?

18. Name several types of balanced apparatus used in radio.

19. Name several types of unbalanced apparatus used in radio.

20. Why may a thin, conducting metal container be used for shielding.

PROBLEMS

1. A telephone line for transmitting a radio program is under test. A voltage of 0.775 volt is impressed on the circuit and 0.001 watt flows into the line. The input current and voltage are approximately in phase. At the distant end a 600-ohm resistor is connected across the line and the power received is 0.0003 watt. Calculate the power loss in decibels, the voltage decrease in decibels, and the current decrease in decibels.

2. As explained on page 14, zero level for sound measurements is 10^{-16} watt. What is the speech power of a +70 decibel sound?

3. On page 30, it states that the output of a typical ribbon microphone is −81 decibels per bar, where zero level is 0.0125 watt. What is the corresponding power output in watts?

4. A vacuum-tube oscillator has an open-circuit voltage output that is substantially constant at 25 volts and has an internal impedance of $Z_s = 600 + j50$ ohms. What is the maximum power that can be drawn from this oscillator, and what should be the impedance of the load that will draw maximum power?

5. For Prob. 4, calculate the total power delivered by the oscillator and the power that is delivered to the load. What is the efficiency of the oscillator?

6. Two coils are inductively coupled, the mutual inductance being 0.02 henry. The coils are identical, their impedances being $Z = 100 + j60$ ohms at 1000 cycles. The secondary coil is connected in series with a load resistance of 40 ohms. A voltage of 10.0 volts at 1000 cycles is impressed across the primary.

a. Calculate the impedance reflected into the primary.

b. Calculate the total primary impedance.

c. Calculate the primary current.

d. Calculate the voltage induced in the secondary.

e. Calculate the current flowing through the load.

7. A certain radio speech-input channel is designed to take the output of a 50-ohm microphone. It becomes necessary to operate a 600-ohm telephone program circuit into this microphone channel. What are your suggestions?

8. A series circuit consists of a 2.0-microhenry inductor and a 100-ohm resistor. Find the parallel circuit that is equivalent at 10^7 cycles.

9. The circuit of Fig. 64 can be used to increase the resistance of a device. If R is 73 ohms and it is to be increased to 500 ohms, what value of L must be used at 15,000,000 cycles. What must be the value of C?

10. The circuit of Fig. 65 can be used to decrease the resistance of a device. If R is 200 ohms and it is to be reduced to 73 ohms, what value of L must be used at 1,000,000 cycles. What must be the value of C?

CHAPTER V

TRANSMISSION LINES, CABLES AND NETWORKS

When radio programs or messages originate at remote locations, the electric signals conveying the programs or messages usually are conducted to the distant radio-transmitting station over audio-frequency lines or cables. (Sometimes, point-to-point radio is used.) At the input and output ends of these lines and cables, and at various points along the circuits, electric networks such as filters, pads, equalizers, and balancing networks are inserted in the transmission circuits.

The audio-frequency signals received at the radio transmitter are impressed on the speech-input circuits, are used to modulate the carrier (page 42), and the resulting radio-frequency signals then are transmitted to the antenna. Often the antenna is located several hundred feet from the transmitting set. Radio-frequency lines and cables are used to transmit the radio-frequency signals from the transmitting set to the antenna.

At the distant receiving station the receiving antenna extracts a small amount of electric energy from the passing radio signal wave. This extracted energy is the received radio signal, and often it is conducted from the receiving antenna to the radio-receiving set over wire lines and cables.

Furthermore, in high-frequency equipment, short sections of radio-frequency transmission lines and cables are used instead of coils and condensers.

From these brief statements it is apparent that those in the field of radio must understand transmission lines, cables, and networks. Accordingly, these subjects will be considered in this chapter.

The Transmission Line.—A simple transmission line consists of two parallel conductors suitably spaced and held in position by insulators. The familiar open-wire telephone line mounted on crossarms attached to poles is an illustration. An antenna feeder that consists of parallel wires separated by ceramic spacers is another.

From the electrical viewpoint, a two-wire transmission line is

far more than just two wires. Each elemental (or small) length of line has capacitance C between the wires, each elemental length of line has inductance L, and each elemental length has resistance R. Also, each elemental length has leakage G, but with good low-loss insulation this often is negligible except at very high radio frequencies. Several elemental sections of a transmission line are shown in Fig. 67. It is very important to note that this diagram applies to two short parallel rods (sometimes called **Lecher wires**) which are commonly used in high-frequency radio. In fact, two parallel rods a few centimeters long constitute a transmission line if the frequency is sufficiently high.

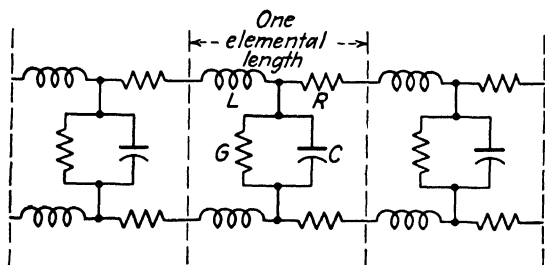



Fig. 67.—Each elemental length of a transmission line consists of inductance L, resistance R, capacitance C, and leakage G.

The transmission line guides electromagnetic energy from the source (such as a radio transmitter) to the load (such as a transmitting antenna). Of course, current flows in the line wires, but such viewpoints are inadequate. *The real purpose of a transmission line is to guide energy from the source to the load.*

The electric energy, constituting the radio (or other) signal to be transmitted, is in the form of an **electromagnetic wave.** An electromagnetic wave consists of electric lines of force and of magnetic lines of force. These electric and magnetic fields as they travel along the transmission line transfer the energy from, for instance, a transmitter to an antenna.

In Fig. 67, if an alternating voltage is impressed on the line, alternating current will flow into the line. The alternating voltage will establish an electric field in the region between the line wires. The alternating current will produce magnetic fields in the region around the wires. This action occurs progressively from section to section. These two fields acting together constitute the electromagnetic wave. The fields are shaped as shown in Fig. 68.

Transmission of Electromagnetic Waves over Wire Lines.—As was shown in the preceding section, when an alternating voltage is impressed on a line, moving electric and magnetic fields are established and these constitute an electromagnetic wave. This wave contains the electric energy to be transmitted. The spark at the switch blades when the current flowing to a coil is interrupted is evidence that a magnetic field contains energy. The spark produced when a charged condenser is short-circuited is evidence that an electric field contains energy.

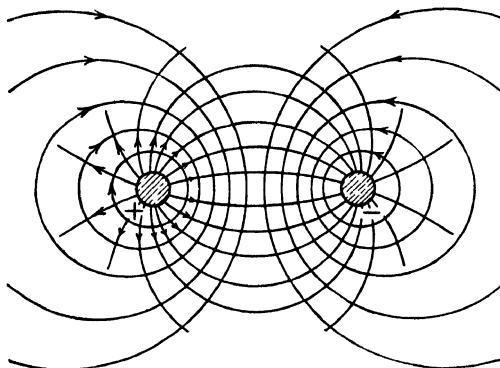



Fig. 68.—The currents in the wires of a transmission line establish magnetic fields around the wires, and the voltage between the wires establishes an electric field between the wires. These two fields contain electric energy and together constitute an electromagnetic wave.

Once the electromagnetic wave is established on the line, the wave continues to travel along the line until the energy contained is dissipated in line losses, or absorbed by a connected load. Thus, unless the line is lossless, the wave is **attenuated** and gradually dies out as it travels along the line.

In elementary electrical theory it is taught that current produces a magnetic field and that the magnetic-field strength is determined by the magnitude of the current. Also, it is taught that voltage produces an electric field and that the electric-field strength is determined by the magnitude of the voltage. These statements are true, but the reverse also is true. Magnetic fields establish currents, and electric fields establish voltages. Thus it is possible and convenient to measure the strength of the magnetic component of an electromagnetic wave traveling along a transmission line by measuring the magnitude of the current in the line. Also, it is pos-

sible and convenient to measure the strength of the electric component of an electromagnetic wave by measuring the magnitude of the voltage between the wires.

The creation and propagation of an electromagnetic wave on a line is shown in Fig. 69. It is assumed that the line extends to infinity. When the impressed alternating voltage increased from *A* to *B*, the electric field (between the wires) and the magnetic field (around the wires) was established on the transmission line and the portion of the electromagnetic wave *A-B* started moving down the line with a velocity approaching the velocity of light

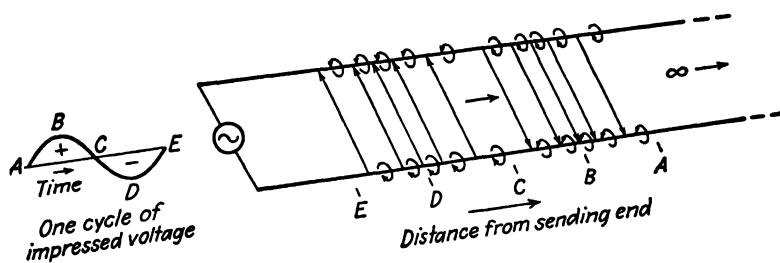



Fig. 69.—When the impressed voltage increased from *A* to *B*, the portion *A-B* of the electromagnetic wave was established on the line and started down the line toward infinity. When the impressed voltage decreased from *B* to *C*, the portion *B-C* of the wave was established, etc. Once created, these waves represent electric energy on the line, and they travel along the line toward infinity. If the line contains loss, the energy is absorbed gradually. The magnetic lines of force around the wires are continuous.

(186,300 miles per second, or 300,000,000 meters per second). As the impressed voltage decreased from *B* to *C*, the portion of the electromagnetic wave *B-C* was established on the line, and started moving down the line. As the voltage built up in the negative direction, portion *C-D* was established, and as the voltage decreased to zero *D-E* was produced. As soon as a field component is established, it starts down the line, so that after *one cycle* of the impressed voltage has occurred, one **wavelength** of the resulting electromagnetic wave has been established along the line as shown.

Because of the difficulty in drawing the magnetic and electric fields, and because these are measured in terms of current and voltage, it is more convenient to deal with currents and voltages. Accordingly, Fig. 70 has been included. In the upper drawing the arrows *along* the wires represent *currents* in the wires. The arrows *between* the wires represent *voltages*. The magnitude or strength of the current is indicated by the density (or closeness) of the arrows

along the wires, and the direction of current flow at each point is indicated by the direction of the arrows. The magnitude or strength of the voltage is indicated by the density (or closeness) of the arrows between the wires, and the direction of the voltage at each point is indicated by the direction of the arrows.

In elementary electrical theory it is taught that the voltage across all parts of a parallel circuit is the same and that the current

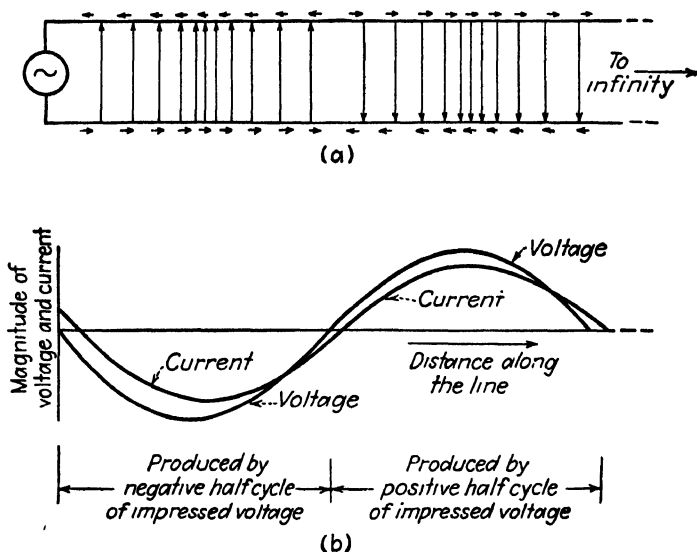



FIG. 70.—Voltage and current distribution along a line at the instant the impressed voltage is zero and starting to increase in a positive direction. Because the characteristic impedance is slightly capacitive (as it is at audio frequencies), the current maximum and minimum values occur ahead of corresponding voltage values. Note that the X-axis represents distance along the line and does not represent time. Values above the line are positive, and those below are negative.

in all parts of a series circuit is the same. That is true for **electrically short circuits** in which the frequency is so low and the line is so short that only a very small fraction of a wavelength can exist on a line. In audio-frequency and radio-frequency transmission lines and cables, **electrically long circuits** are under consideration. A large fraction of a wavelength, and sometimes many wavelengths, may exist on such circuits. The simple concepts applied to electrically short circuits do not hold. At the ultrahigh radio frequencies two wires but a few centimeters long constitute an electrically long circuit.

The fact that in Fig. 70 the currents are flowing toward each

other and away from each other, and are different in the same wire, and the fact that the voltages at the various points are different in magnitude and direction can be accounted for in this way: The various portions of the electromagnetic wave are bundles of electric-signal energy traveling along the line. These were created by the transmitting source at different instants, and each bundle of signal energy is independent and remains independent of the others. It also is convenient to picture that as the various bundles of electromagnetic energy travel along the line they cause motions of the electrons in the wire. Thus one bundle of energy at one point may cause currents in one direction, but at another point the moving electromagnetic wave may be causing electron motion (current flow) in another direction. This viewpoint may be applied to the voltages between the wires. The energy bundles cause excess and deficiencies in electrons along the wires, and an electric field exists between such portions.

The lower diagram in Fig. 70 indicates a further simplification in drawing electromagnetic waves. Here the direction of the magnetic component of the electromagnetic wave (and the direction of the current) is represented by whether the line is above (positive) or below (negative) the axis. The strength of the magnetic field (and the magnitude of the current) is represented by the height of the line above the zero axis. Corresponding statements apply to the electric components and the voltages. In Fig. 70*b* the corresponding values of the waves are not exactly at the same points at a given instant because the input impedance of the line is assumed to be slightly capacitive (page 129).

Characteristic Impedance.—The equivalent circuit of a two-wire transmission line or simple two-wire cable is shown in Fig. 67. If a voltage is impressed across the input terminals (for example, at the left of Fig. 67), the voltage will force a current into the line.

The magnitude of the input current and the phase angle between the impressed voltage and the input current will be determined by R, L, C, and G of Fig. 67. Thus the size and material of the wires, the spacing of the wires, the insulators used, etc., will determine the magnitude and phase angle of the input current.

If the impressed voltage is divided by the input current of *any* line, a value of impedance is obtained, and for *any* line this is called the **input impedance.** For the lines which have been considered in Figs. 69 and 70, and *which extended to infinity,* the input im-

pedance is called the **characteristic impedance.** Thus the characteristic impedance, usually designated by Z_O, *is the input impedance of a line infinite in length.* The characteristic impedance of a line is give by the equation

$$Z_O = \sqrt{\frac{R + j2\pi fL}{G + j2\pi fC}}. \tag{49}$$

When R is the combined resistance in ohms of the two wires per unit length *of line*, L is the series inductance in henrys of the two

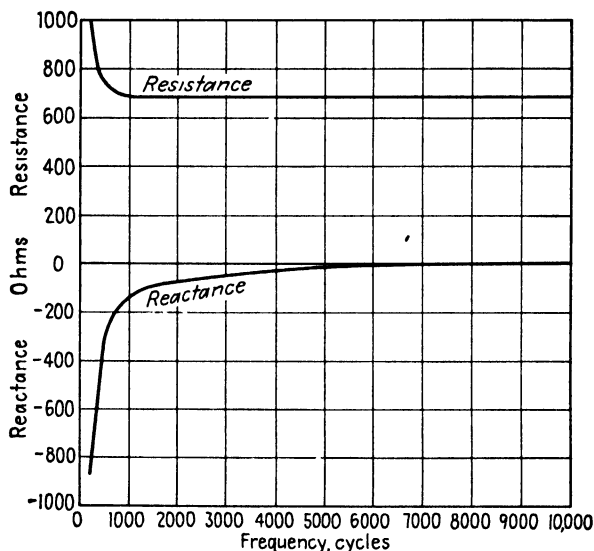



FIG. 71.—Resistance and reactance (capacitive) components of the characteristic impedance of a typical audio-frequency transmission line.

wires per unit length of line, G is the shunt leakage in ohms per unit length, C is the capacitance in farads between wires per unit length, and f is the frequency in cycles per second, then Z_O will be the characteristic impedance in both magnitude (ohms) and angle. Any convenient unit of length can be used, provided that it is the same for R, L, G, and C. From Eq. (49) it is seen that the characteristic impedance of a line is different at various frequencies. The discussions just given, and Eq. (49), apply to a two-wire audio-frequency telephone line or to a two-wire radio-frequency transmission line. The characteristic impedance of a line for audio-frequency signals is shown in Fig. 71. Note that the characteristic

impedance contains a capacitive reactance component that is fairly high at low frequencies and that decreases toward zero at the higher audio-frequency values. The characteristic impedance of a radio-frequency transmission line will be considered on page 144.

Line Terminations.—The transmission lines considered in Figs. 69 and 70 extended to infinity. For this reason an electromagnetic wave, once established on the line, will travel down the line forever. The electric and magnetic components of the wave would be attenuated, or decreased in amplitude, as they traveled along the line because of the series resistance losses R and the shunt leakage losses G of Fig. 67.

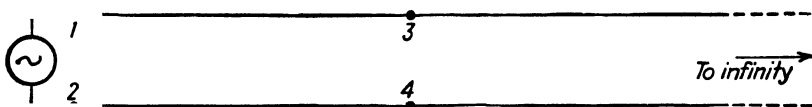



FIG. 72.—An infinite line. If the line section between points 1-2 and 3-4 is removed, the line remains infinite in length. Thus, input impedance measurements made at points 1-2 (with the generator disconnected) will be the same as the input impedance measured at points 3-4 (with the line to the left disconnected). This input impedance is the characteristic impedance Z_O.

Of course in practice a line is finite in length and is used to transmit electric signal energy from one point to another. The important point is this: How should a line be terminated at the distant receiving end so that it will abstract, or take from the line, all the signal energy that arrives? It is apparent that there are an infinite number of impedance values that the terminating equipment may have, and the question arises: What is the correct terminating value for the maximum abstraction of energy?

To answer these questions, refer to Fig. 72. By definition the input impedance at points 1–2 is the characteristic impedance (the line being infinite in length). Suppose now that a finite section of the line, the length between points 1–2 and 3–4, is removed, and that the input impedance of the remainder of line (extending from point 3–4 to infinity) is measured. This measured impedance value *will also be the characteristic impedance*, because removing a finite section of line (between points 1–2 and 3–4) from an *infinite* line does not appreciably alter its length.

Again considering the infinite line of Fig. 72, when an electromagnetic wave travels down the line, *all* the energy that arrives at point 3–4 will enter the section extending from point 3–4 to infinity. There is no discontinuity at point 3–4 in the wave path, and the

energy flows on into section 3–4 to infinity, which has an input impedance of Z_O, the characteristic impedance of the line.

Now suppose that the section 3–4 to infinity is disconnected and that a resistor and a reactor in series are connected across the line at point 3–4. If the resistor and a capacitive reactor are of the same magnitudes as the resistive component and the reactive component of the characteristic impedance Z_O, then the wave will not know whether it is entering some more of the line or the resistor and reactor termination. The energy in the wave will accordingly enter the termination and will be dissipated in the resistor. Of course there would be little object in heating a resistor, and in practice the termination would be the input impedance of the equipment that the signal transmitted by the line is to drive, properly transformed (page 109) to equal the characteristic impedance of the line.

Transmission Equations.—As has been explained, when an electromagnetic wave travels along a two-wire line (or cable), the wave is attenuated or weakened. Because the line is composed of resistance R, inductance L, capacitance C, and leakage G, as shown in Fig. 67, a finite time is required for the wave to travel a given distance. Inductance opposes the rise of current, and capacitance slows the rise of voltage; thus in a sense the wave progresses from section to section down the line. The maximum velocity it can have is that of light (page 126).

The effect of the line constants R, L, C, and G in attenuating the wave, and in regulating its velocity, is expressed by the **propagation constant** γ of the line. The equation for the propagation constant for a two-wire line or two-wire cable circuit is

$$\gamma = \sqrt{(R + j2\pi fL)(G + j2\pi fC)}. \tag{50}$$

When the values are substituted in this equation, the value of γ will contain two terms. One will be a "real" number, and the other will be an "imaginary" number. The real number is the term not preceded by j, and the imaginary number is the number that is preceded by j. Thus, the propagation constant may be thought of as having two terms

$$\gamma = \alpha + j\beta, \tag{51}$$

where α is the **attenuation constant** and determines the rate at which the wave is attenuated and β is the **phase constant** and de-

termines how rapidly the wave travels. The velocity is given by the equation

$$V = \frac{2\pi f}{\beta} \tag{52}$$

The units involved and the use of Eqs. (49), (50), (51), and (52) will be explained in an illustrative problem (page 134).

Constants of an Open-wire Line at Audio Frequencies.—As was shown in Fig. 67, a transmission line is composed of resistance, inductance, capacitance, and leakage. The numerical values of these are called the **line constants** or **line parameters.** Each of these will be considered now.

Resistance R.—This constant includes all *series* losses, that is, the I^2R losses caused by current flow and the associated magnetic field. Such losses are the ordinary direct-current resistance of the wires; the skin-effect losses caused by the alternating currents crowding to the surfaces of the wires, thus reducing the effective cross-sectional conducting area; the eddy current losses; and the magnetic hysteresis losses. Hard-drawn copper wires usually are employed for transmission lines, and these have a resistance of about 1.03 times that of annealed-copper wires. The resistance of the wires used for transmission line *at audio frequencies* can be calculated by Eq. (14), page 50, following the method there explained. For the *two* wires of a transmission line 1 mile long, using hard-drawn wires 0.165 inch in diameter, the resistance at 20° C. will be (neglecting skin effect, eddy current, and hysteresis losses, as is common practice at audio frequencies)

$$R = \frac{10.7 \times 2 \times 5280}{(165)^2} = 4.15 \text{ ohms.}$$

Inductance L.—The series self-inductance of the two wires of a transmission line *at audio frequencies* is approximately

$$L = 0.001481 \log_{10} \frac{b}{a} + 0.00016 \text{ henry per mile of line,} \tag{53}$$

where b is the distance between the wires measured from center to center, and usually expressed in inches in telephone work, and a is the radius of each wire, usually in inches. Of course other units can be used, provided that they are the same for each dimension. Lines for transmitting audio frequencies often are spaced 12 inches

apart. Thus the self-inductance of an audio-frequency line 1 mile long composed of 0.165-inch wires spaced 12 inches apart is

$$L = 0.001481 \log_{10}\left(\frac{12}{0.0825}\right) + 0.00016 = (0.001481 \log_{10} 145.5)$$
$$+ 0.00016 = (0.001481 \times 2.163) + 0.00016$$
$$= 0.00321 + 0.00016 = 0.00337 \text{ henry.}$$

Capacitance C.—The capacitance between the two wires of a transmission line is approximately

$$C = \frac{0.01941 \times 10^{-6}}{\log_{10}(b/a)} \text{ farad per mile of line,} \qquad (54)$$

where b and a are as previously considered. Thus the capacitance of a transmission line 1 mile long composed of two wires 0.165 inch in diameter and 12 inches apart is

$$C = \frac{0.01941 \times 10^{-6}}{\log_{10}(12/0.0825)} = \frac{0.01941 \times 10^{-6}}{\log_{10} 145.5} = \frac{0.01941 \times 10^{-6}}{2.163}$$
$$= 0.00898 \text{ microfarad.}$$

Leakage G.—This is not the simple direct-current leakage between wires. The leakage G is a dissipative factor that includes all the *shunt* (or parallel) losses in the transmission line, that is, the E^2G losses caused by the voltage and the electric field between wires. It includes the direct-current leakage losses, and also losses caused by the alternating electric field in the crossarms, insulators, poles, etc. These losses are so complicated that they are not computed as for R, L, and C, but are measured experimentally. Often the leakage losses are neglected.

Telephone lines often are used for transmitting the audio-frequency signals of messages and programs that are to be transmitted by radio systems. Standard open-wire telephone lines are of hard-drawn copper of three sizes: wires 0.165 inch in diameter, wires 0.128 inch in diameter, and wires 0.104 inch in diameter. These commonly are spaced 12 inches part. Their constants vary with temperature and frequency, and are influenced by the presence of other wires on the same poles and crossarms. Constants for the three standard sizes of wires are given in Table VII. This table is for 1000 cycles, 20°C., dry-weather conditions, and for a two-wire hard-drawn copper circuit on a 40-wire pole line. The data may be used for an isolated two-wire line with little error. It will be noted that the wire sizes do not follow the A.W.G.

system. The constants can be calculated as explained in this section for wire sizes other than those given. It will be noted that these values, which are for an actual line, agree closely with the calculated values.

TABLE VII.—CONSTANTS FOR OPEN-WIRE AUDIO-FREQUENCY TRANSMISSION LINES
(Per mile of line)

Wire diameter, inches	Resistance, ohms	Inductance, henrys	Capacitance, microfarads	Leakage, micromhos	Characteristic impedance, ohms	Loss, decibels
0.165	4 11	0.00337	0.00915	0.29	$612 \angle -5.35$	0.0300
0.128	6.74	0.00353	0 00871	0 29	$650 \angle -8\ 32$	0.0462
0.104	10 15	0.00366	0 00837	0 29	$692 \angle -11.75$	0.0660

Transmission over an Open-wire Line at Audio Frequencies.—The use of the equations for characteristic impedance, propagation constant, and the data just given will be demonstrated by numerical calculations.

Illustrative Problem.—A radio-transmitting station is located 18 miles from a remote program pickup point. It is possible to use an open-wire transmission line for feeding programs from the remote studio to the transmitter. It is decided to use standard telephone-line construction, and 0.165 inch hard-drawn copper wires. A complete analysis of the transmission characteristics at 1000 cycles is desired. The values of the line constants from Table VII instead of the calculated values will be used.

Solution.—Step 1. Calculate the characteristic impedance, using Eq. (49).

$R + j2\pi fL = (4.11 + j6.283 \times 1000 \times 0.00337) = (4.11 + j21.2) = 21.6 \angle 79^\circ$ ohms.

$G + j2\pi fC = (0.29 \times 10^{-6} + j6.283 \times 1000 \times 0.00915 \times 10^{-6}) = (0.29 \times 10^{-6} + j57.5 \times 10^{-6}) = 57.5 \times 10^{-6} \angle 90^\circ$ mho, the in-phase component being negligible.

Inserting these values in Eq. (49) gives

$$Zo = \sqrt{\frac{21.6 \angle 79^\circ}{57.5 \times 10^{-6} \angle 90^\circ}} = \sqrt{372{,}000 \angle -11^\circ} = 610 \angle -5.5^\circ,$$ a value that checks with Table VII.

Step 2. Calculate the propagation constant, using Eq. (50). The values previously calculated can be used. Thus Eq. (50) becomes

$\gamma = \sqrt{(21.6 \angle 79^\circ)(57.5 \times 10^{-6} \angle 90^\circ)} = \sqrt{0.001242 \angle 169^\circ} = 0.0353 \angle 84.5^\circ = 0.0353 \cos 84.5^\circ + j0.0353 \sin 84.5^\circ = 0.0353 \times 0.0958 + j0.0353 \times 0.9954 = 0.00339 + j0.0352.$

Step 3. Calculate the line loss in decibels per mile. This is done from the first term computed for γ. The value of 0.00339 is in **nepers** per mile. To convert from nepers to decibels the factor 8.686 is used. Thus the loss in decibels is $0.00339 \times 8.686 = 0.0295$ decibels per mile. This agrees closely with Table VII.

Step 4. Calculate the wave velocity and the wavelength. This is done from the second term computed for γ. The wave velocity is, from Eq. (52), $V = 6.283 \times 1000/0.0352 = 179{,}000$ miles per second, or slightly less than that of light. Wavelength is equal to wave velocity divided by frequency, and for the line under consideration is $\lambda = 179{,}000/1000 = 179$ miles.

Step 5. After the line is constructed, it is decided to check its operation by impressing 1.0 volt at 1000 cycles on the sending end and measuring the input current and power, and the output current and power. For this check the line will be terminated in its characteristic impedance, although from a practical standpoint a 600-ohm resistor would be satisfactory.

The current input will be $I = E/Z = 1.0/610 = 0.00164$ ampere or 1.64 milliampere, and this current will lead the voltage by 5.5°.

The power entering the line will be $P = EI \cos \theta = 1.0 \times 0.00164 \times 0.9954 = 0.001633$ watt or 1.633 milliwatts.

The power at the distant end can be computed from a rearrangement of Eq. (27). From logarithms, if $n = 10 \log_{10} P_1/P_2$, then $P_2 = P_1/(10^{0.1 \times n})$, where P_2 is the power output, P_1 is the power input, and n is the loss of the line in decibels. For the 18-mile line under consideration the loss is $18 \times 0.03 = 0.54$ decibel. This value being used, the power that reaches the distant termination will be $P_2 = 1.633/(10^{0.1 \times 0.54}) = 1.633/10^{0.054} = 1.633/1.132 = 1.44$ milliwatts.

The voltage at the distant end can be computed from the received power and the resistance. Thus, $P = E^2/R = E^2/(Z \cos \theta)$, and $1.44 \times 10^{-3} = E^2/(610 \times 0.9954)$. From this, $E = 0.937$ volt.

The current at the distant end will be $I = E/Z = 0.937/610 = 0.001535$ ampere, or 1.535 milliamperes.

In transmitting a message or program over an open-wire line the signal voltage impressed on the line might be considerably greater than the 1.0 volt considered here. An open-wire line may pick up noise from adjacent electric-power lines, and a high signal level often is used to ensure that the signal strength always is much greater than the noise; that is, that the **signal-to-noise** level is high.

The line just considered should be balanced with respect to ground (page 116) and with respect to other paralleling circuits (if any) by transposing the positions of the conductors at intervals of perhaps 1000 feet, depending on conditions. This equalizes the capacitance to ground and maintains a high degree of balance. Transpositions are effective in reducing noise pickup from parallel-

ing power lines and in keeping crosstalk to adjacent telephone circuits (if any) at a low level.

Transmission over Cables at Audio Frequencies.—The ordinary cable consisting of twisted pairs of copper wires in a lead sheath operates fundamentally as an open-wire line. The twisting maintains the balance to ground (sheath) and reduces crosstalk. The equivalent circuit of Fig. 67 applies. The line constants are much different than for an open-wire line. Because it is desired to have the cable diameter small, and because usually it is desired to have hundreds of pairs in a telephone cable, the wires are small and close together. Small wires may be used because they do not have to support themselves as for open-wire lines. Annealed-copper wires are used, and these follow A.W.G. numbers.

The resistance R is much larger than for open-wire lines. It can be found by Eq. (14), page 50. The capacitance C is very much larger than for open-wire lines, because the wires of a cable are close together and have paper insulation with a dielectric constant greater than unity. The self-inductance L is very low compared with an open-wire line, because the wires of a cable are so close together that the tendency of the current in one wire to produce a magnetic field is almost completely offset by the magnetic-field producing tendency of the same current in the other wire of the pair. The leakage G is higher than for dry open-wire lines because the cable wires are so close together; also, there are dielectric losses caused by the electric field in the paper insulation. However, the leakage is very constant because the cables are sealed from moisture, and the cable leakage is less than for open-wire lines when these lines are exposed to fog, etc. Equations (53) and (54) do not hold for cables, because the wires are very close together.

Cable circuits used for audio-frequency transmission often are **inductively loaded.** Inductance in the form of small coils of special design is inserted at regular intervals in each wire of a cable pair. In the United States, open-wire lines are not loaded. The main purpose of loading is to reduce the loss (or attenuation) in the cable, although other benefits also are derived from loading. Audio-frequency long-distance or toll cable pairs usually are loaded, but cable pairs in an exchange or city area seldom are loaded, unless it is necessary for special reasons. A cable has a characteristic im-

pedance and propagation constant and follows the same general transmission theory as does an open-wire line.

It does not appear necessary to give detailed methods of transmission calculations for audio-frequency cables. Cables would be purchased, not constructed on the job, as would an open-wire line. If the cable loss in decibels and the characteristic impedance are known, then calculations similar to those of the last part of the illustration in the preceding section can be made. Pertinent data for loaded and nonloaded cables are given in Table VIII.

TABLE VIII.—CHARACTERISTICS OF CERTAIN AUDIO-FREQUENCY TELEPHONE CABLES

Type, cables	Size wire, A.W.G.	Type, loading	Characteristic impedance, ohms	Attenuation, decibels per mile	Cutoff frequency
Toll.	16	None	326 $\angle -40°40'$	0 72	
Toll.	16	B–88–S	1541 $\angle -1°29'$	0.16	5600
Toll.	16	B–22–N	803 $\angle -4°39'$	0 24	11,000
Exchange. . .	26	None	1007 $\angle 44°30'$	2 67	
Exchange . .	22	None	576 $\angle 43°48'$	1 79	
Exchange. . .	19	None	402 $\angle 42°48'$	1 26	

The data given are for standard Bell System telephone cables. The loading designations, such as B–88–S, specify the inductance to be added, etc. The circuits considered are two-wire circuits (not grounded telephone circuits, or phantom telephone circuits). Note that loading causes a cable to exhibit a cutoff frequency (page 165) beyond which the cable will not transmit. Thus, the B–88–S loading would cause the cable to cut off at 5600 cycles, and therefore would not permit wide-band transmission. On the other hand, the B–22–N loading would permit wide-band transmisison up to 11,000 cycles. A typical problem involving a cable circuit will be considered.

Illustrative Problem.—A radio-transmitting station is located near the edge of a large city and close to the route of a toll cable. It is decided to obtain audio-frequency channels from the telephone company over a 16-gauge cable pair with B–22–N loading. The signal will be fed in at the studio at +6 volume units. It is 14.7 miles from the studio to the transmitter location. What will be the received signal strength at the transmitter?

Solution.—In reference to Table VIII, the total loss will be $14.7 \times 0.24 = 3.53$ decibels. The received volume level will be $+6 - 3.53 = +2.5$ volume units (approximately). If this is more than is required by the speech-input

equipment at the transmitter, the signal volume can be reduced by suitable means (page 163).

Constants of Open-wire Lines at Radio Frequencies.—The equations for the constants R, L, C, and G of Fig. 67, page 124, usually are given in different form for radio-frequency lines than for the audio-frequency lines previously considered. For this reason, the values at radio frequencies will now be considered. The equations are given in many forms and degrees of exactness. An attempt has been made to select the most simple, yet reliable, forms. These equations assume that *the spacing between wires divided by wire diameter is at least about 8.*

Resistance R.—**Skin effect** becomes very great at radio frequencies. Because of the high frequency, the current is crowded to the surface layers of the conductors. For this reason, a hollow tube is as good a conductor at radio frequencies as is a solid rod of the same outside diameter. At ultrahigh radio frequencies, silver plating a tube reduces the resistance. This is because silver is an excellent conductor and most of the current flows near the surface. The resistance at radio frequencies of hard-drawn copper wires can be computed by the equation

$$R = \frac{8.4\sqrt{f}}{a}, \tag{55}$$

where R is the resistance in microhms of *both* wires of a radio-frequency open-wire transmission line 1 meter long, f is the frequency in cycles per second, and a is the radius of the wires in centimeters.

Inductance L.—The series self-inductance of a radio-frequency open-wire transmission line is

$$L = 0.921 \log_{10} \frac{b}{a}, \tag{56}$$

where L is the inductance in microhenrys of both wires of a line 1 meter long, b is the distance between wire centers, and a is the radius of the wires. These often are measured in centimeters in radio work, but can be in other units if the same unit is used for each.

Capacitance C.—The capacitance between the wires of a radio-frequency open-wire transmission line is

$$C = \frac{0.000012}{\log_{10}\frac{b}{a}}, \tag{57}$$

where C is the capacitance in microfarads per meter and b and a are as discussed for inductance.

Leakage G.—For radio frequencies this factor becomes very difficult to evaluate. No data appear to be readily available. Of course there would be some loss caused by the voltage and the electric field between wires. For making calculations on radio-frequency transmission lines this loss often is neglected, although sometimes it is quite important.

Transmission Equations for Radio-frequency Open-wire Lines.—The general expression for the characteristic impedance of a transmission line was given by Eq. (49), and the general expression for the propagation constant (from which the attenuation and wave velocity are obtained) is given by Eq. (50). These equations are used at *audio frequencies*. For a line at *radio frequencies*, these general equations may be written in simpler forms,[1] which will be discussed.

The characteristic impedance of an open-wire line at radio frequencies is

$$Z_O = \sqrt{\frac{L}{C}} = 276 \log_{10} \frac{b}{a}, \tag{58}$$

where Z_O is the characteristic impedance in ohms when L is the inductance in henrys and C is the capacitance in farads for a given length of line. The characteristic impedance is a value of pure resistance because Eq. (58) contains no reactive or j term. This equation is derived from Eq. (49) where at radio frequencies the R and G terms become negligible. This also is shown by Fig. 71, which indicates that as the frequency is increased the reactive term of the characteristic impedance approaches zero. In the second form of Eq. (58), b is the distance between wire centers and a is the radius of the wires, both expressed in the same units.

The attenuation constant α of a line at radio frequencies, when the leakage G is neglected, is

$$\alpha = \frac{R}{2\,Z_O}, \tag{59}$$

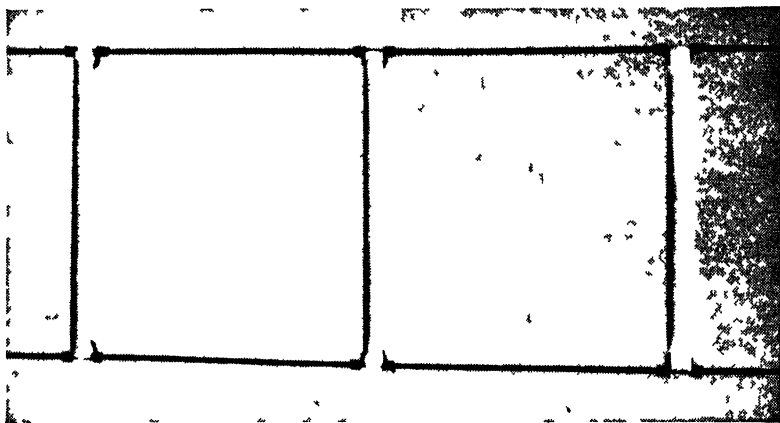
where α will be in nepers and must be multiplied by 8.686 to change to decibels, R is the resistance as given by Eq. (55), and Z_O is as

[1] Fleming, J. A., "The Propagation of Electric Currents in Telephone and Telegraph Conductors," D. Van Nostrand Company, Inc.

given by Eq (58) If R is expressed in ohms per meter, then α will be the attenuation per meter

The velocity of a radio-frequency electromagnetic wave on an open-wire radio transmission line is approximately that of light or about 300,000,000 meters per second The wavelength is approximately

$$\lambda = 3 \times 10^8 f, \qquad (60)$$

where λ is in meters when f is in cycles For an open-wire line at radio frequencies the calculated values of λ sometimes are reduced

Photograph of a short section of a radio frequency transmission line The conductors are of copper wire and the spreaders are of low loss ceramic material (*E F Johnson Co*)

by a factor of 0 975 to give more exact results, because the theoretical and actual wavelength often are slightly different (see also page 524)

Transmission over an Open-wire Line at Radio Frequencies.—A numerical solution will be used to illustrate the application of the preceding information to a radio-frequency transmission line

Illustrative Problem —The transmitting antenna is located 550 feet from the transmitting set, and it is to be connected with an open-wire line that is composed of two No 4 A W G hard-drawn copper wires spaced 18 inches apart and carefully constructed and insulated so that the leakage is negligible It is desired to transmit 3 0 kilowatts input at 3,200,000 cycles The antenna is so matched to the line that the termination is essentially the characteristic impedance of the line A complete analysis is desired

Solution —Step 1 Convert all measurements to meters and centimeters Line length = 550 × 0 3048 = 167 5 meters Spacing between wires =

18 × 2.54 = 45.7 centimeters. Diameter of No. 4 A.W.G. wire is 0.204 inch, and radius = 0.204 × 2.54/2 = 0.258 centimeter.

Step 2. Calculate the resistance, using Eq. (55).

$$R = \frac{8.4\sqrt{f}}{a} = 8.4\sqrt{3.2 \times 10^6}/0.258 = 58{,}300 \text{ microhms per meter}$$
$$= 0.0583 \text{ ohm per meter.}$$

Step 3. Calculate the characteristic impedance, using Eq. (58).

$$Z_O = 276 \log_{10}\frac{b}{a} = 276 \log_{10}\frac{45.7}{0.258} = 276 \log_{10} 177$$
$$= 276 \times 2.248 = 620 \text{ ohms.}$$

Step 4. Calculate the attenuation constant, using Eq. (59), and the total line loss. $\alpha = R/2Z_O = 0.0583/2 \times 620 = 0.000047$ neper per meter = 0.000047 × 8.686 = 0.000408 decibel per meter. Total loss = 0.000408 × 167.5 = 0.0684 decibel.

Step 5. With 3.0 kilowatts input, the power delivered to the antenna would be (using the method outlined on page 135, Step 5) $P_2 = P_1/10^{0.1 \times n})$ $= 3/(10^{0.1 \times 00.684}) = 3/(10^{0.00684}) = 3/1.0159 = 2.950$ kilowatts. This means that 50 watts would be lost in transmission, and that the efficiency of the line would be 2.95/3.0 = 0.985 or 98.5 per cent.

Step 6. Calculate the impressed voltage and input current. When Z_O is a pure resistance, as it is for the high-frequency line, $P = E^2/Z_O$, and $E = \sqrt{PZ_O} = \sqrt{3000 \times 620} = \sqrt{1{,}860{,}000} = 1365$ volts. $P = EI$, and $I = P/E = 3000/1365 = 2.2$ amperes.

Radiation losses have been neglected in the line calculations just made because they are negligible except at very high frequencies. Attention is called again to the fact that these equations and calculations apply to an open-wire radio-frequency line when the spacing between wires divided by the wire diameter is about 8 or more. When very large conductors, such as rods or tubes, are used, or when wires are twisted together for providing a transmission line, the equations must be modified. In general, such lines have very low characteristic impedance, and transmit at low voltages and high currents. For example, the characteristic impedance of two parallel tubes 0.5 inch in outside diameter and spaced 5 inches apart is about 350 ohms, as shown in Fig. 73. The characteristic impedance of a twisted pair of No. 14 wire, such as is sometimes used in house wiring, is about 70 ohms. Parallel tubes and twisted wires are transmission lines in the strict sense, but are seldom used for transmitting power over distances exceeding perhaps 100 feet.

Radio-frequency Transmission over Coaxial Cables.—A cross-sectional diagram of a coaxial cable is shown in Fig. 74. Often it consists of a copper tube with a copper conductor held at the

center, the tube being one "side" of the circuit and the central conductor the other "side." The conductor at the center is held in place by insulating spacers placed at regular intervals. Sometimes solid insulating material, such as polyethylene, is used for small flexible coaxial cables.

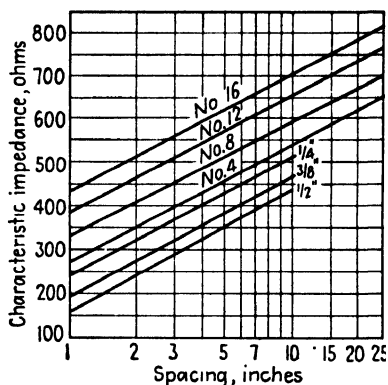



FIG. 73.—Characteristic impedances of transmission lines composed of copper wires and tubes. Spacings are center to center. "No. 16," etc., refer to A.W.G. wire sizes. Dimensions in inches are outside diameters of tubes.

Coaxial cables are used very extensively, particularly for important installations, to connect radio-transmitting sets to transmitting antennas located at a distance. Such cables also are used to connect radio-receiving antennas to radio-receiving sets. The present discussion will be confined to coaxial cables in transmitting.

There are several reasons for coaxial cables being used extensively, among them are the following: (*a*) The radio-frequency energy is entirely confined within the cable, and radiation is zero; (*b*) the cable has a low characteristic impedance, and this sometimes is advantageous; (*c*) it is an unbalanced circuit (as contrasted with the two-wire line previously considered which is a balanced circuit), and sometimes an unbalanced transmission circuit is more desirable; (*d*) last, and probably most important, the coaxial cable can be buried in the ground where it is not likely to be injured, and where it offers no obstructions. An underground cable is susceptible to electrolysis and to corrosion, and cables often are damaged in these ways. Usually pressure tanks of dry nitrogen gas are attached to coaxial cables to keep the interior dry. Should a small leak occur, nitrogen gas flows out and prevents the entrance of moisture. Also, the drop in pressure, as noted on the pressure gauge, indicates that trouble has developed.

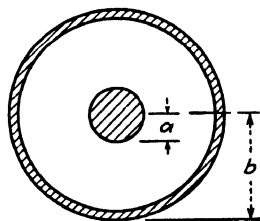



FIG. 74.—Cross section of a coaxial cable, such as is used in radio.

Constants of Coaxial Cables at Radio Frequencies.—As for the open-wire line, from an electrical viewpoint a coaxial cable is more

than just a wire at the center of a tube. The coaxial cable is, electrically, a network similar to that of Fig. 67, except that it is completely unbalanced with respect to ground because the outer conductor is grounded.

Resistance R.—For the two conductors of a coaxial cable at radio frequencies the resistance is given by the equation

$$R = 4.2\sqrt{f}\left(\frac{1}{a} + \frac{1}{b}\right), \tag{61}$$

where R is in microhms per meter of cable when f is in cycles, a is the radius of the inner conductor in centimeters, and b is the *inside* radius of the outer conductor in centimeters.

Inductance L.—The series self-inductance of a coaxial cable is

$$L = 0.46 \log_{10} \frac{b}{a}, \tag{62}$$

where L is the inductance in microhenrys per meter of cable, when b and a are as previously explained.

Capacitance C.—The capacitance between the conductors of a coaxial cable is

$$C = \frac{0.0000241\ K}{\log_{10}(b/a)}, \tag{63}$$

where C is microfarads per meter of cable, when K is the dielectric constant of the insulating medium between the two conductors, and b and a are as previously explained. From a practical standpoint the dielectric constant of air or nitrogen is unity. In coaxial cables using disks or spacers to hold the inner conductor at the center of the outer sheath, the equivalent dielectric constant would be slightly different from unity, but little error is involved if the insulator spacings are of several inches or more. For solid dielectrics the value of K would be the dielectric constant of the material, which for polyethylene is about 2.25 up to at least 10^8 cycles.

Leakage G.—Because the coaxial cables are sealed from moisture the ordinary current leakage is low. Since in many cables air or nitrogen is the dielectric, the losses caused by the alternating radio-frequency electric field is low. Of course part of the electric field also penetrates the spacers and a loss is caused in the spacers. For this reason spacers that offer low loss at radio frequencies are used. Equations are available for correcting for the effect of the spacers, but in most practical work these losses are neglected. Where solid dielectrics are used this loss may become very im-

portant, and particularly at high radio frequencies may be the factor determining the suitability of the cable. Such losses are difficult to evaluate.

Transmission Equations for Radio-frequency Coaxial Cables.—Because the coaxial cable is an electrical network similar to Fig. 67, the coaxial cable obeys the general transmission-line theory given in the preceding pages and has a characteristic impedance and propagation constant that determine its performance.

The characteristic impedance of a coaxial cable at radio frequencies is

$$Z_O = \sqrt{\frac{L}{C}} = \sqrt{\frac{0.46 \log_{10}(b/a)}{\frac{0.0000241\,K}{\log_{10}(b/a)}}} = \frac{138}{\sqrt{K}} \log_{10} \frac{b}{a}, \qquad (64)$$

where K, b, and a are as explained in the preceding section. The characteristic impedance of a coaxial cable at radio frequencies is essentially pure resistance.

The attenuation constant of a coaxial cable with air or nitrogen dielectric and at radio frequencies is

$$\alpha = \frac{R}{2Z_O}, \qquad (65)$$

where α will be in nepers when R, from Eq. (61), and Z_O, from Eq (64), are both in ohms. The loss in decibels equals the loss in nepers multiplied by 8.686.

The velocity of propagation of a radio-frequency electromagnetic wave on a coaxial cable with air or nitrogen dielectric is that of light approximately, or about 300,000,000 meters per second. This neglects the effect of the spacers. The theoretical wavelength would be as given by Eq. (60). Where a solid dielectric is used, the velocity and wavelength would be less than for air by the factor $1/\sqrt{K}$, where K is the dielectric constant of the dielectric material. Because of the factors neglected, the actual wave velocity and wavelength might be about 0.9 times the theoretical value.

It can be shown[1] theoretically and verified experimentally that for least attenuation the ratio of b/a should be 3.6. Many coaxial cables are constructed with this ratio of dimensions, irrespective of the over-all sizes. For this ratio, Eq. (64) gives the characteristic impedance of a coaxial cable with air dielectric to be Z_O = 138

[1] Sterba, E. J., and C. B. Feldman, Transmission Lines for Short-wave Radio Systems, *Proceedings of the Institute of Radio Engineers*, Vol. 20, No. 7, July, 1932.

$\log_{10} 3.6 = 138 \times 0.5563 = 77$ ohms approximately. For this reason, coaxial cables often are termed low-impedance circuits. There appears to be a growing tendency to use ratios other than 3.6. Thus coaxial cables with characteristic impedances lower than 77 ohms are now available. Cables having a characteristic impedance of 52 ohms often are used.

Transmission over a Coaxial Cable at Radio Frequencies.—The use of the preceding theory and equations will be illustrated by a practical problem.

Illustrative Problem.—A coaxial cable is to be used to connect a 250-watt radio transmitter operating at 60 megacycles with the transmitting antenna that is 500 feet away. A coaxial cable having an inner radius of the outer conductor equal to 0.5 inch and a ratio of b/a equal to 3.6 is available. The cable will be filled with nitrogen. The antenna is matched to the cable. A complete analysis of the problem is desired in order to ascertain if this cable will be satisfactory for practical operation.

Solution.—Step 1. Convert dimensions b and a (Fig. 74) to centimeters and express the distance in meters. $b = 0.5 \times 2.54 = 1.27$ centimeters. $a = b/3.6 = 1.27/3.6 = 0.353$ centimeters. Distance $= 500 \times 0.3048 =$ 152.4 meters.

Step 2. Calculate the resistance, using Eq. (61).

$R = 4.2\sqrt{f}\left(\frac{1}{a} + \frac{1}{b}\right) = 4.2\sqrt{60 \times 10^6}\left(\frac{1}{0.353} + \frac{1}{1.27}\right) = 4.2 \times 7.75$ $\times 10^3(2.83 + 0.788) = 4.2 \times 7.75 \times 3.618 \times 10^3 = 117{,}700$ microhms per meter $= 0.1177$ ohm per meter.

Step 3. Calculate the characteristic impedance, using Eq. (64).

$Z_O = 138 \log_{10} \frac{b}{a} = 138 \log_{10} \frac{1.27}{0.353} = 138 \log_{10} 3.6 = 138 \times 0.5563 =$ 77 ohms.

Step 4. Calculate the attenuation constant, using Eq. (65), and the total line loss. $\alpha = R/(2Z_O) = 0.1177/(2 \times 77) = 0.000764$ neper per meter $=$ $0.000764 \times 8.686 = 0.00664$ decibel per meter. Total loss $= 0.00664 \times$ $152.4 = 1.012$ decibels.

Step 5. With 250 watts input to the cable, the power delivered to the antenna would be (page 135, Step 5) $P_2 = P_1/(10^{0.1 \times n}) = 250/(10^{0.1 \times 1.012})$ $= 250/(10^{0.1012}) = 250/1.262 = 198$ watts. This means that 52 watts would be lost in transmission, and that the efficiency of the cable would be $198/250 = 0.793$ or 79.3 per cent.

Step 6. Calculate the impressed voltage and input current. When Z_O is essentially pure resistance as for the coaxial cable, $P = E^2/Z_O$, and $E = \sqrt{PZ_O} = \sqrt{250 \times 77} = \sqrt{19{,}250} = 139$ volts. $P = EI$, and $I =$ $P/E = 250/139 = 1.8$ amperes.

Wave Reflection on Lines and Cables.—The lines and cables of the preceding pages were used for transmitting between two points

electric signal energy in the form of electromagnetic waves. For this purpose the lines and cables were terminated in their characteristic impedance. Under these conditions, all the energy arriving at the distant end was absorbed by the termination, or terminal, equipment.

If a line or cable is not terminated in its characteristic impedance, then all the electric energy in the **initial wave** that arrives at the receiving end is not absorbed. When the electromagnetic wave containing the electric signal energy reaches the termination, the mismatched impedance (differing from Z_O) acts as an irregularity in the transmission path offered to the wave, and only *part* of the energy is absorbed. From the conservation of energy principle, the part that is not absorbed (or lost in some other way, such as by radiation) must flow back toward the sending end. This energy also is in the form of an electromagnetic wave, and the reflected energy constitutes a **reflected wave.** If a line or cable is open-circuited at the distant end, then the "termination" of infinite impedance can absorb no power, because the current through it would be zero, and *all* the energy that reaches the distant end must be reflected back toward the sending end. If a line or cable is short-circuited at the distant end, then the termination of zero impedance can absorb no power, because no voltage can exist across it, and *all* the energy that reaches the distant end must be reflected back toward the sending end.

To digress for a moment from lines and cables, it is of interest to compare the above phenomena with the action of an inductor or capacitor. As mentioned on pages 45 and 46, inductors and capacitors absorb energy for a part of the cycle of the impressed alternating voltage, and return much of the energy to the source during the next part of the cycle. Coils and condensers that have no losses return *all* the energy, and are said to be reactances because of these phenomena.

In this respect, open-circuited and short-circuited transmission lines and cables act like coils and condensers. They alternately accept power as an initial wave from the source, and return it as a reflected wave to the source. If a line or cable is open-circuited or short-circuited, if it has no losses and does not radiate power, then these open or shorted lines or cables would behave like inductive or capacitive reactances, and would be equivalent to inductors and capacitors.

In radio, particularly at high radio frequencies, advantage is taken of these phenomena, and open-circuited and short-circuited lines and cables are used extensively instead of coils and condensers. At first thought this would seem very awkward, but it is not so at all. For the very high radio frequencies surprisingly short lengths of lines or cables can be used instead of coils and condensers.

The sections immediately following will consider the subject of wave reflection, and will explain in more detail why open-circuited and short-circuited lines or cables can be used as reactive circuit elements, and for other useful purposes.

Voltage Distribution on an Open-circuited Lossless Line.—As mentioned in the preceding pages, sections of open-wire lines and cables are used as inductive and capacitive elements in radio. They also are used for impedance transformers, and for other purposes. To explain their action, wave-reflection phenomena must be studied. To simplify the matter, the line will be assumed to be *lossless and cause no attenuation.* Although this is not quite possible, two short silver-plated rods or tubes approach this condition. The discussion in this section will be limited to voltage distribution on a lossless line.

Electromagnetic waves were studied on page 125, and as Fig. 70*b* indicated, the voltage at various points along the line at a given instant could be indicated by a curve drawn with reference to a zero axis that represented the line. When the curve was above the line, it meant that the voltage was positive at that point. When the curve was below the line, it meant that the voltage was negative at that point. The height of the curve above (or below) the line indicated the magnitude of the voltage at that point.

Throughout this chapter stress has been placed on the fact that an electromagnetic wave contains energy, and for this reason, when an electromagnetic wave traveling along a line strikes an open circuit or a short circuit, the energy must be reflected back toward the sending end. Since an electromagnetic wave is composed of electric and magnetic fields and their accompanying voltages and currents, and since voltages and currents, not fields, ordinarily are measured, the statement can be made that when a voltage wave strikes an open circuit or short circuit the voltage wave is reflected back toward the sending end.

This situation is depicted in Fig. 75. The line under considera-

tion is a lossless line 5/4 λ in length, where $\lambda = 3 \times 10^8/f$ meters and f is the frequency of the applied voltage in cycles. Figure 75*a* indicates by a solid line the distribution of the initial wave at the

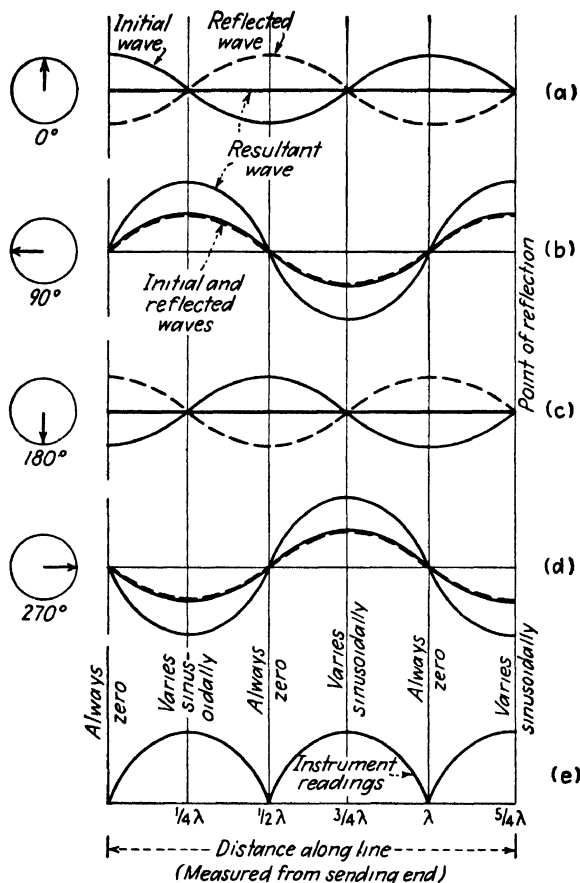



FIG. 75.—Curves of voltage relations on an open-circuited lossless line or current relations on a short-circuited lossless line. At certain points along the line the voltage or current *always* is zero. At all other points the magnitude of voltage or current is as indicated by curve (*e*) and marked "instrument readings." The voltage or current at other than the zero points is the familiar alternating voltage or current and has the frequency of the voltage impressed at the sending end. (*Adapted from "Communication Engineering," by W. L. Everitt, McGraw-Hill Book Company, Inc.*)

instant that the voltage wave impressed at the sending end is a positive maximum, as the vector in the small circle at the left indicates. By using the value at the distance λ from the sending end as an example, the distribution of the initial voltage wave can

be verified in this way: As Fig. 75*a* shows, the voltage wave at a distance λ (one wavelength) from the sending end is a maximum positive value *at the instant under consideration.* In the time the wave traveled the distance λ, the impressed voltage varied through 360°, or one cycle. Thus, the portion of the wave *now* at a distance λ from the sending end is a positive maximum value, because 360° ago the impressed voltage was a positive maximum. The voltage *now* (meaning at the instant under consideration, as shown in Fig. 75*a*) at a distance λ was sent out 360° ago. By following the same method, the initial voltage wave can be located at any point and at any of the four positions (or instants of consideration) of the impressed voltage vector.

When a voltage wave strikes an open at the distant end of a line, it merely flows back just as it would have gone on. The electric field extends between positive electric charges on one wire and negative electric charges on the other wire, and since these charges cannot flow across the open, the *direction of action* or sign of the voltage wave is not changed at the point of reflection (but of course its *direction of travel is changed,* and the wave flows back toward the sending end). The same method of reasoning given in the preceding paragraph can be used to locate the reflected voltage wave.

The initial voltage wave flowing out toward the point of reflection, and the reflected voltage wave flowing back, combine at each point along the line to give a resultant wave. For Fig. 75*a, at the instant depicted* (when the impressed voltage wave is a maximum positive value) the resultant voltage is zero all along the line. At the instant that the impressed voltage is as shown in Fig. 75*b*, the initial and reflected waves add at each point along the line. At the instant that the impressed voltage wave is a negative maximum (Fig. 75*c*), the initial and reflected waves cancel and produce zero voltage at all points along the line. For the position of the impressed voltage shown in Fig. 75*d* the initial and reflected waves add.

The positions of the initial and reflected waves, and the resultant waves, have been shown in Fig. 75 for one complete cycle. If a voltmeter is connected across the line at any point, it will read the resultant voltage. The readings of a voltmeter connected at the various points along the line are shown in Fig. 75*e*. This curve shows that for a lossless, open-circuited line, 5/4 wavelength long, the voltage as read by a voltmeter would be (theoretically) zero

at the sending end; a maximum $\frac{1}{4}\lambda$ from the sending end; zero at $\frac{1}{2}\lambda$; maximum at $\frac{3}{4}\lambda$, zero at λ, and maximum at $\frac{5}{4}\lambda$. This is a peculiar phenomenon, but it exists nevertheless. When a voltage is impressed on an open-circuited lossless line, there are *points* along the line where the voltage is maximum and *points* where it approaches zero. This phenomenon has been called a **standing wave,** or more correctly (according to the Standard Definitions of Electrical Terms) a **stationary wave.** The term standing wave is used widely in radio. These are poorly chosen terms: What is called a stationary wave, or a standing wave, really is not a wave in the usual sense at all; what is called a stationary wave, or a standing voltage wave, is really a plot or a curve showing the magnitude of the voltage that would be indicated by a voltmeter connected at various points along the line. Furthermore, this would be an alternating voltage fundamentally the same as encountered along any energized line.

Voltage Distribution on a Short-circuited Lossless Line.—When an electromagnetic wave strikes the distant end of a short-circuited line, the voltage component must be zero because no voltage can exist across a short circuit of zero impedance. The electric charges (between which the electric lines of force extend or exist) pass from one wire to the other through the short circuit, and at the instant of reflection the voltage component is reversed in direction. For this reason the reflected voltage wave does not travel back toward the sending end as it does for an open-circuited line. Instead, the reflected component is reversed 180° at the instant of reflection, and then travels back toward the sending end. This leads to a distribution of the initial and reflected components, and to the resultant component, as shown in Fig. 76. The various positions of the initial and reflected waves can be explained as in the preceding section, keeping in mind that the voltage wave is reflected with a 180° shift in phase at the short circuit.

A voltmeter connected at the various points along the line would read the values shown in Fig. 76*e*. Thus, when an alternating sine-wave voltage is impressed on a lossless short-circuited line, the distribution of voltage along the line is as indicated.

Current Distribution on an Open-circuited Lossless Line.—As was shown in Fig. 70*b*, a curve can be used to indicate the magnitude of the current along the line, and if the characteristic impedance of a line is essentially pure resistance, as it is in radio, the

initial current and the initial voltage waves would be in phase. (They are slightly out of phase in Fig. 70*b*, because the discussion there was for audio frequencies where the characteristic impedance contains some capacitive reactance.)

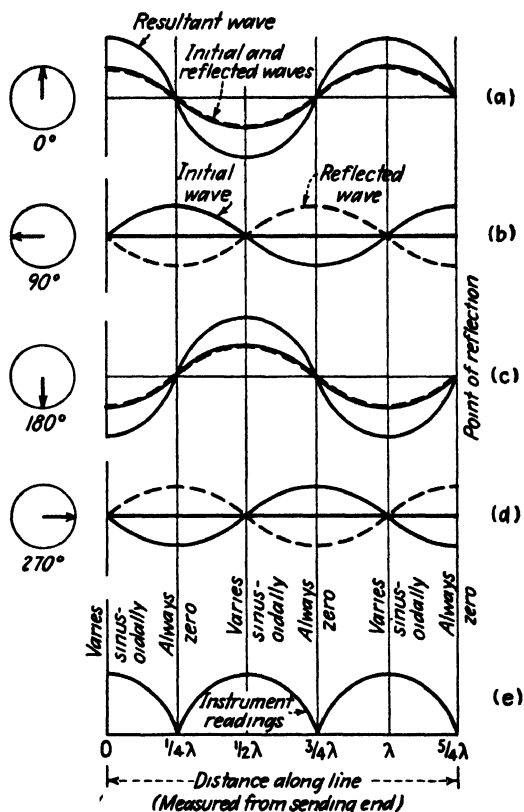



FIG. 76.—Curves of current relations on an open-circuited lossless line or voltage relations on a short-circuited lossless line. At certain points along the line the current or voltage always is zero. At all other points the current or voltage is as indicated by curve (*e*), and marked "instrument readings." The current or voltage at other than the zero points is the familiar alternating current or voltage and has the frequency of the voltage impressed at the sending end. (*Adapted from "Communication Engineering," by W. L. Everitt, McGraw-Hill Book Company, Inc.*)

The same method as previously outlined can be applied to current waves. Thus (in Fig. 76*b*) the initial current wave arriving at the end of the line is a positive maximum because it is the value that was created $5/4\lambda$, or 450° before the instant under consideration. Turning the impressed voltage vector backward 450°, or $1\frac{1}{4}$ cycle,

places it in the position to create a positive maximum current value.

The reflected current wave can be located by similar reasoning, but the fact must be considered that at the instant of reflection from an open circuit the current is reversed 180°. This reversal is because the charges that arrive at an open circuit must reverse and flow back, thus constituting a reversal of 180° for the current and magnetic component of the electromagnetic wave.

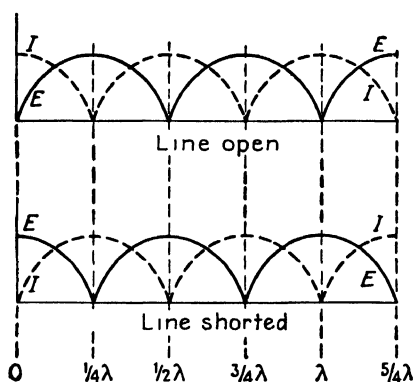



FIG. 77.—Current and voltage relations on open- and short-circuited lines having no attenuation. These are *not* initial and reflected waves. They are the (*e*) curves of Figs. 75 and 76. They are plots of the voltages and currents at various points along the line.

Because of these reversals, the reflected current waves and the resultant current waves are as indicated in Fig. 76. Thus the current that would be indicated by an ammeter would approach zero at certain points along the line, and would be maximum at others. Furthermore, this would be fundamentally the same as any alternating current. The curve (Fig. 76*e*) of current values at various points along the line is called in radio a "standing current wave."

Current Distribution on a Short-circuited Lossless Line.—The same method is used to determine this distribution. However, at the instant of reflection there is no 180° shift in phase of the current. The reason is that the short circuit provides a path for the charges constituting the current flow, and there is no reversal in direction, such as occurs at an open circuit. For this reason, Fig. 75 applies. The distribution of current, as it would be read by ammeters at various points along the line, is as shown in Fig. 75*c*.

Current and Voltage Distributions on Lossless Lines.—The relations discussed in the four preceding sections, and as shown in Figs. 75 and 76, are summarized in Fig. 77. These curves show that where the voltage values are maximum the current values are minimum, and that for the lossless line they approach zero as a limiting value.

Now of course, these explanations were for a line 5/4λ in length.

The question arises: What will be the distribution along a line other than $\frac{5}{4}\lambda$ in length? The answer to this is simple. At the distant end of an open line the current must be zero, and because of the reciprocal relation between current and voltage, the voltage will always be maximum. At the distant end of a short-circuited line,

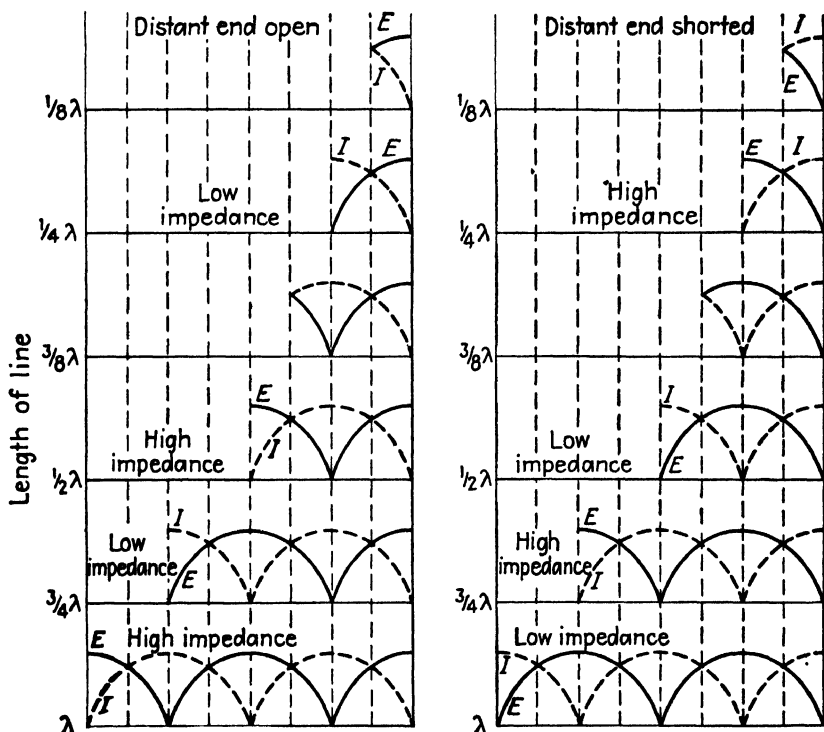



FIG. 78.—The current and voltage relations on open-circuited and short-circuited lossless lines of various lengths are as shown. For an open-circuited line of any length, the voltage at the open end is maximum, and the current is zero. For a short-circuited line of any length, the voltage at the shorted end is zero, and the current is maximum.

the voltage must be zero and the current will always be maximum. *The terminal conditions at the distant end, not the sending end, always determine the voltage and current distributions along the line.* Because of this, in investigating conditions along any line, the study is made starting at the distant end. Thus the voltage and current distribution along lines of various length are shown in Fig. 78. This is, of course, derived from Fig. 77.

Input Impedance of Open-circuited and Short-circuited Lossless Lines.—As has been stressed, these lines cannot absorb power because no attenuation exists, and because when the wave strikes the distant end all the energy that was initially sent into the line is reflected back toward the sending end.

Now a transmission line (or a cable) has two terminals to which the generator driving the line is connected. For the generator not to be able to supply power over a given period of time one of the following conditions must exist: (*a*) The input impedance to the line must be a zero value of pure resistance; (*b*) the input impedance to the line must be an infinite value of pure resistance; (*c*) the input impedance must be a value of pure capacitive reactance; or (*d*) the input impedance must be a value of pure inductive reactance. These conclusions are based on the fact that $P = EI \cos \theta$.

The input impedance of open-circuited and short-circuited lines and cables have the values just listed, the specific numerical value depending on the length of line, the frequency, the wave velocity, and the termination. That these statements are true can be proved by vector diagrams.

Open-circuited Line Less than $\frac{1}{4}\lambda$ *in Length.*—A vector diagram for studying the input impedances of an open-circuited line is shown in Fig. 79. To study the input impedance, the relations between the initial and reflected voltages and currents at the sending end must be studied. The line is assumed for convenience to be slightly less than one-fourth wavelength long. Since the distant end of the line fixes conditions, the voltage E_O at the distant open end is taken as the reference on the X axis. The initial voltage E_i at the sending end is slightly less than 90° ahead of E_O, as indicated by the angle α, because E_i is a voltage impressed on the line *later* than was E_O. Because the characteristic impedance is pure resistance, the initial current I_i is in phase with E_O. Likewise, the current I_O just arriving at the open end of the line is the phase with voltage E_O. When the current is reflected from the open end, however, it is reversed in direction by the 180° phase shift, and hence becomes $-I_O$.

Figure 79 represents conditions on a line at a given instant, that is, at the instant E_O has just arrived. Thus the reflected component of voltage that has arrived back at the sending end at this instant is E_R It has the vector position shown because it was sent out on the line before E_O. The reflected component of the

current that has arrived back at the sending end is α degrees behind the position $-I_O$. The actual voltage at the sending end E_S is the vector sum or resultant of E_i and E_R, and has the position indicated. The actual current at the sending end I_S is the resultant of I_R and I_i. Thus the sending-end current I_S leads the sending-end voltage E_S by 90°, and, hence, the input impedance of a lossless open-circuited line less than one-fourth wavelength long is a value

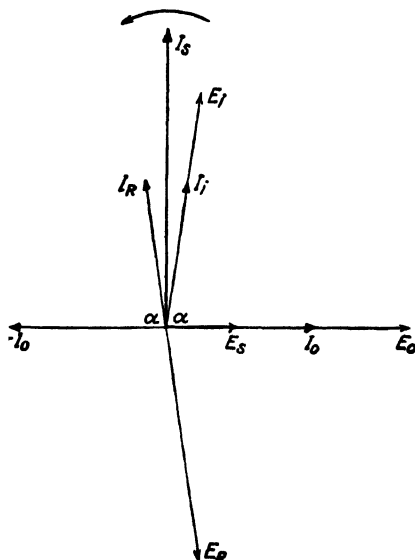



FIG. 79.—Vector diagram for an open-circuited lossless line less than $\frac{1}{4}\lambda$ in length. For such a line the input impedance is pure capacitive reactance as shown by the fact that the sending-end current I_S leads the sending-end voltage E_S by 90°.

of pure capacitive reactance and the line is equivalent to a capacitor.

Open-circuited Line Greater than $\frac{1}{4}\lambda$ in Length.—These relations are shown in Fig. 80. Because the line length is greater than $\frac{1}{4}\lambda$, the angles corresponding to α of Fig. 79 will be greater than 90°. When this vector diagram is studied, it is found that the sending-end current *lags* the sending-end voltage by 90°, and, hence, the input impedance of a lossless open-circuited line greater than one-fourth wavelength long is a value of pure inductive reactance and the line is equivalent to an inductor. This holds so long as the length is greater than $\frac{1}{4}\lambda$ but less that $\frac{1}{2}\lambda$.

Short-circuited Line Less than $\frac{1}{4}\lambda$ in Length.—A vector diagram

similar to those just shown will indicate that the input impedance of a short-circuited line less than one-fourth wavelength long will be a value of pure inductive reactance and that the line is equivalent to an inductor.

Short-circuited Line Greater than ¼λ in Length.—Vector diagrams will show that the input impedance of this line is a value of pure capacitive reactance and that the line is equivalent to a capacitor. This holds so long as the length is greater than ¼λ but less than ½λ.

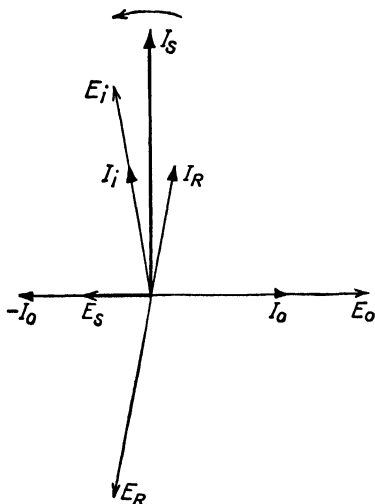



FIG. 80.—Vector diagram for an open-circuited lossless line greater than ¼λ in length. For such a line the input impedance is pure inductive reactance as proved by fact that the sending-end current I_S lags the sending-end voltage E_S by 90°.

Input Impedance of Any Line or Cable.—In radio, the open-circuited and short-circuited (essentially) lossless lines are called **resonant lines** because the input impedances of these lines vary in a manner similar to the input impedances of ordinary resonant circuits composed of inductance and capacitance.

It is possible to write an equation that will give the equivalent input impedance of any length of any line or cable terminated in any manner, but this equation is quite involved. For the radio-frequency lossless line the equation is much simplified into the following forms:

$$Z_{OC} = Z_O \cot \frac{2\pi X}{\lambda} \angle -90^\circ \quad \text{and} \quad Z_{SC} = Z_O \tan \frac{2\pi X}{\lambda} \angle +90^\circ, \tag{66}$$

where Z_{OC} is the input impedance (a pure capacitive reactance as the $\angle -90^\circ$ indicates) of an open-circuited lossless line *less than one-quarter wavelength long*, Z_{SC} is the input impedance (a pure inductive reactance as the $\angle +90^\circ$ indicates) of a short-circuited lossless line *less than one-quarter wavelength long*, Z_O is the characteristic impedance of the line (pure resistance at radio frequencies), λ is the wavelength (equal to wave velocity divided by frequency, page 140), and X is the length of the section measured in the same units as λ.

By using different values of X between zero and $\frac{1}{4}\lambda$, reactances from zero to infinity are available theoretically.

Illustrative Problem.—Two parallel silver-plated brass tubes have a characteristic impedance of 350 ohms, and each is 1 meter long. They are driven at one end with a frequency of 60 megacycles. Calculate the input impedance to the tubes when the distant end is open-circuited and when it is short-circuited.

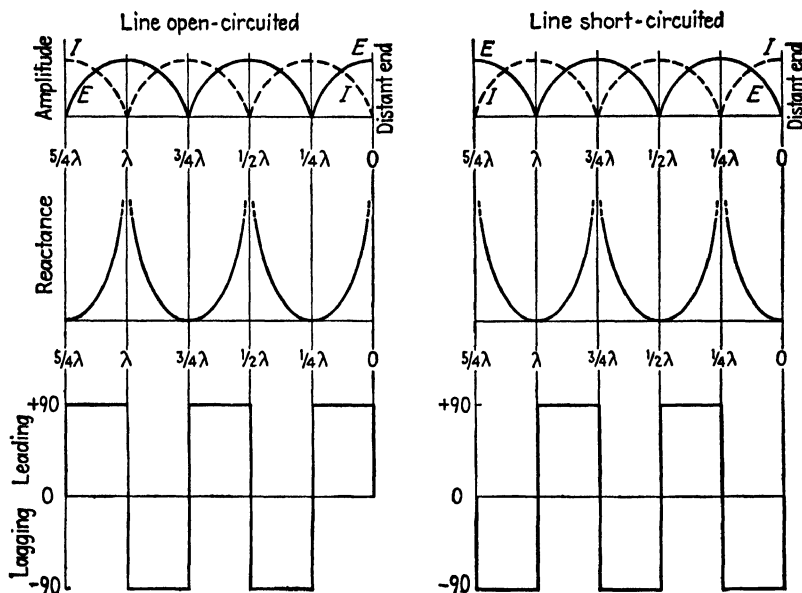



FIG. 81.—Input-impedance characteristics of open-circuited and short-circuited lossless lines of various lengths, as measured from right to left. For instance, for the open-circuited line, if the length is less than $\frac{1}{4}\lambda$, the input impedance will be pure capacitive reactance (leading 90°). If the length is less than $\frac{1}{2}\lambda$, but greater than $\frac{1}{4}\lambda$, the input impedance will be pure inductive reactance (lagging 90°).

Solution.—Step 1. Evaluate $2\pi X/\lambda$ of Eq. (66). $\lambda = V/f = 3 \times 10^8/6 \times 10^7 = 5$ meters. $2\pi X/\lambda = 6.28 \times 1.0/5 = 1.257$ radians $= 1.257 \times 57.3 = 72°$.

Step 2. Calculate Z_{OC}, the input impedance to the open line, using Eq. (66). $Z_{OC} = Z_O \cot 72° = 350 \times 0.3249 = 113.7 \underline{/-90°}$ ohms. This would be equivalent to a condenser having a capacitance of $X_C = 1/2\pi fC$, $C = 1/2\pi fX = 1/(6.28 \times 6 \times 10^7 \times 113.7) = 23.3 \times 10^{-12}$ farad.

Step 3. Calculate Z_{SC}, the input impedance to the shorted line, using Eq. (66). $Z_{SC} = Z_O \tan 72° = 350 \times 3.0776 = 1075\underline{/+90°}$ ohms. This would be equivalent to a coil having an inductance of $X_L = 2\pi fL$, $L = X_L/2\pi f = 1075/(6.28 \times 6 \times 10^7) = 2.85 \times 10^{-6}$ henry.

The input impedances (which of course are equal to the impressed voltage divided by the input current) of lossless lines are summarized in Fig. 81. To find the nature of the input impedance of any line either open-circuited or short-circuited, measure off from the right-hand end of the proper diagram the length of the line in question; then, the curves give the input-impedance characteristics of the section. Thus the input impedance of an open-circuited line $\frac{1}{8}\lambda$ in length is a low value of pure capacitive reac-

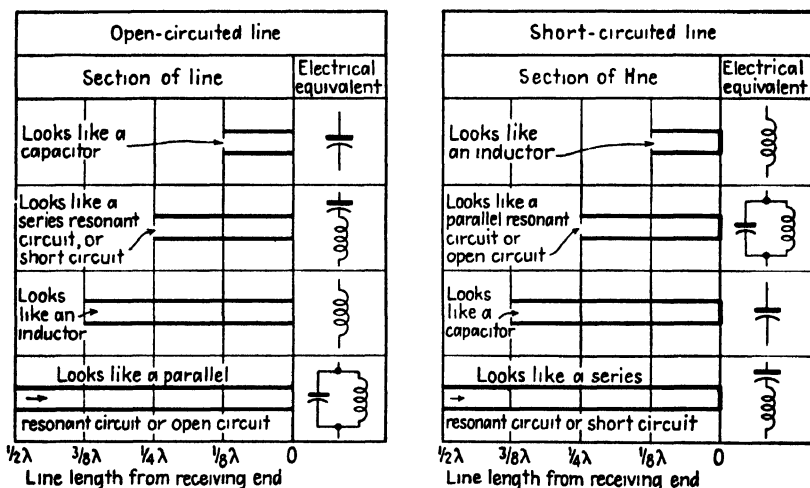



FIG. 82.—These illustrations show the input-impedance characteristics of various transmission-line sections. The equivalent electrical circuits also are shown.

tance. On the other hand, the input impedance of a short-circuited line $\frac{1}{8}\lambda$ in length is a low value of pure inductive reactance. The important input characteristics of resonant lines are summarized in Fig. 82.

Resonant Lines as Transformers.—It has been shown that resonant lines exhibit the characteristics of inductors and capacitors. Referring back to Fig. 78, and also studying Fig. 81, will show that they also exhibit the characteristics of transformers. Thus the input impedance of a lossless line that is $\frac{1}{4}\lambda$ in length and is open at the distant end is zero. Thus, in a sense, an infinite impedance has been transformed into a zero impedance. These figures show also that if the line is short-circuited, then the input impedance is infinite. Hence, a zero impedance has been transformed by the line into an infinite impedance.

Sometimes it is desired to transform a zero impedance into an infinite impedance, an example being shown in Fig. 83. Here it is desired to hold a high-frequency radio-transmission line above some supporting structure. This can be done with a continuous metal bracket, if the bracket is $\frac{1}{4}\lambda$ in length. The "short circuit" at the distant end appears as an infinite impedance where the bracket is attached to the high-frequency line, and hence negligible power is

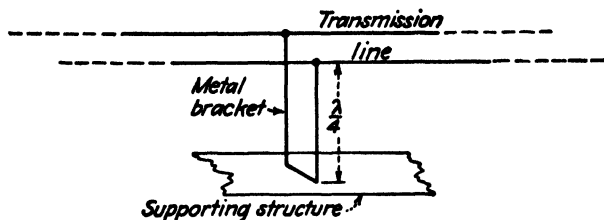



Fig. 83.—A quarter-wave metal bracket can be used as an insulator to support a radio-frequency transmission line. This principle is based on the fact that the input impedance of a quarter-wave shorted section approaches infinity.

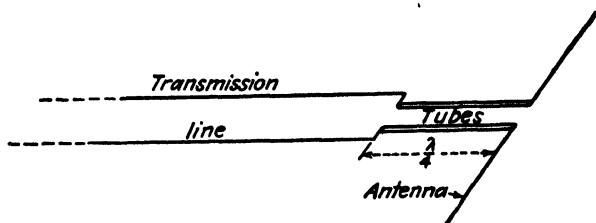



Fig. 84.—Two quarter-wave tubes (or wires) can be used to match the impedance of an antenna to a radio-frequency transmission line.

lost in the metal bracket. Thus a "conductor" becomes an "insulator."

Of course there are many times when it is desired to transform impedances which are not zero or infinity, but which are of finite values. The ways of doing this with resonant lines are many and varied. One simple and very useful application is shown in Fig. 84. Suppose that an antenna which has an input driving impedance of 72 ohms pure resistance is to be matched to a transmission line which has a characteristic impedance of 550 ohms. To do this, a **quarter-wave matching section** is used. This is composed of tubes $\frac{1}{4}\lambda$ long, and of such diameter and spacing that their characteristic impedance is 200 ohms. This is determined by the formula

$$Z_{\text{section}} = \sqrt{Z_{\text{line}}\, Z_{\text{load}}}, \qquad (67)$$

where $Z_{section}$ is the characteristic impedance that the quarter-wave matching section must have in order to match the antenna to the line for maximum power transfer. Thus, $Z_{section} = \sqrt{550 \times 72} =$ 200 ohms (approximately). The dimensions and spacing for wires and tubes to give a certain characteristic impedance are shown in Fig. 73, page 142.

Reflection on Terminated Lines and Cables having Attenuation. —In the preceding pages of this chapter the discussions were confined to a large extent to two types of lines: (*a*) those lines (and cables) which have attenuation and which were terminated in their characteristic impedance because they were to transmit energy; and (*b*) those lines (and cables) which have negligible losses and which were open-circuited or short-circuited at the distant end so that they would reflect energy and act like capacitive or inductive elements. The preceding section considered a circuit terminated in an impedance different from its characteristic impedance and used as an impedance transformer.

The distribution of the current and voltage along a line terminated in its characteristic impedance follows a logarithmic decrease as indicated in the calculations given (see Step 5, page 135). Aplot of line voltage against distance from the source would be a smooth curve for a line which had attenuation and which was terminated in its characteristic impedance.

The distribution of current and voltage (standing wave) along a line not terminated in its characteristic impedance varies between maximum and minimum values. This is because of interaction between the initial and reflected waves. Examples were shown in Figs. 77, 78, and 81 of the distribution for open- or short-circuited lossless lines.

If the line contains attenuation, or if the line is not open- or short-circuited, but is terminated in some impedance other than its characteristic impedance, then the standing wave values will not rise to as high a maximum or fall to as low a minimum. If a transmission line is lossless but has any resistance in its termination, then energy will flow into the line, and the four statements on page 154 will not hold. If a line has attenuation but is open- or short-circuited, then again energy will flow into the line, and these four statements do not hold. When a line has resistance in its termination, or has attenuation, then the input impedance contains a resistance component, and the minimum values of the

current or voltage distributions do not fall to the theoretical zero value.

Because in radio the transmission lines are so short and the losses are so low, the following discussion will be limited to a consideration of a lossless line with a *pure resistance* termination that is different from the characteristic impedance.

Reflection on Terminated Lines and Cables Having No Attenuation.—When a lossless radio-frequency line is not terminated in its characteristic impedance, the magnitude of the so-called "standing wave" is determined by the degree of mismatch. There is therefore a definite relationship between the degree of mismatch and the ratio of the maximum to minimum value of the standing wave. This relation is

$$\text{Standing-wave ratio} = \frac{\text{maximum voltage}}{\text{minimum voltage}} = \frac{Z_O}{Z_R} \quad \text{or} \quad \frac{Z_R}{Z_O} \qquad (68)$$

Thus if the standing-wave ratio is 3 on a line having a characteristic impedance of 500 ohms, the terminating resistance Z_R is either 1500 ohms or 167 ohms. To determine which it is, the standing-wave pattern (that is, the distribution of the voltage along the line) must be studied. If the terminating resistance is greater than Z_O, then the voltage across the terminating resistance will be greater than when the line is terminated in Z_O. If the terminating resistance is less than Z_O, then the voltage across the resistance will be less than when the line is terminated in Z_O.

The preceding discussions have applied only to resistance terminations because they are most common, and because when reactances are considered the problem becomes involved. The discussion has used voltage ratios as illustrations. Of course current ratios can be used as well.

For indicating the presence of standing *voltage* waves along the line, a small neon lamp often is used. When this lamp is held between the wires, the electric field caused by the voltage between the wires causes the gas to glow where the voltage is high, and is insufficient to cause it to glow where the voltage is low. Of course this would only locate the position of the **antinodes** (maximum voltage points) and **nodes** (minimum voltage points), and would not give data for determining voltage ratios. Quantitative data for calculating voltage ratios can be measured in several ways. One of the simplest is to connect a thermocouple milliammeter at the end of a quarter-wave section. Because the resistance of the

thermocouple heater is low, the input impedance at the other end of the quarter-wave section will be very high. This combination then can be connected across a line, as shown in Fig. 85, and used as a high-impedance voltmeter.

For indicating the presence of standing *current* waves on a line a small flashlight bulb often is used. A coil of a few turns of wire is attached to the terminals of the bulb, and the coil and bulb then are held close to the line. The resultant magnetic field caused by the currents in the wires induces sufficient energy in the coil to cause the light to burn. This locates only antinodes and nodes.

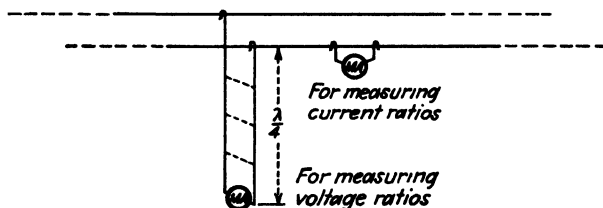



FIG. 85.—For measuring voltage ratios on a line the thermomilliammeter at the end of a quarter-wave line is convenient. Because the instrument has low resistance, the input impedance to the quarter-wave line will approach infinity. For measuring current ratios, there is sufficient IZ drop in each section spanned by the thermomilliammeter leads to cause a current reading.

Quantitative data for determining current ratios can be obtained with a thermocouple milliammeter, as indicated in Fig. 85. Two hooks of stout copper wire are attached to the milliammeter terminals and the combination is hung on one wire of the line and moved along. There is sufficient IZ drop in the length of wire spanned to cause current flow through the milliammeter, and thus it deflects and measures the strength of the current I in the wire.

Networks.—This is a rather general term used in a variety of ways. It will be used here in the sense that a network is a circuit, or device, having two input and two output terminals. Such networks often are inserted between a transmission line, or cable, and the associated load impedance. This is done for several purposes, sometimes to cause a loss, sometimes to compensate for line distortion.

Networks are used in audio-frequency circuits and in radio-frequency circuits for purposes such as the ones just explained. Their use in audio-frequency circuits is so very common that the discussion will be confined largely to networks of the audio-frequency type.

Pads and Attenuators.—A pad is a network composed of resistors and designed to insert a *fixed* loss into a circuit. If the resistors vary little with frequency, a pad attenuates equally over a wide frequency range. Pads often are used in audio-frequency program circuits in this way: Programs are often transmitted at high volume levels to ensure that the signal-to-noise ratio is high. The program then is received at a level too high to impress on the speech-input equipment of the radio transmitter. Pads are used to reduce the volume level to the correct amount.

If a pad is to be placed at the end of a two-wire transmission line or cable between the line (or cable) and the terminating equipment, then (*a*) the pad must be balanced (page 116) just as is the line or cable, (*b*) the pad must have (ordinarily) the same characteristic impedance as the line or cable, and (*c*) the pad must insert the proper loss into the line.

FIG. 86.—A simple pad for inserting a known loss in a circuit. This is a balanced circuit with the same impedance in each side.

If an unbalanced pad were connected to an open-wire line that was exposed to inductive interference from adjacent power lines, the pad would unbalance the circuit and probably cause it to be noisy.

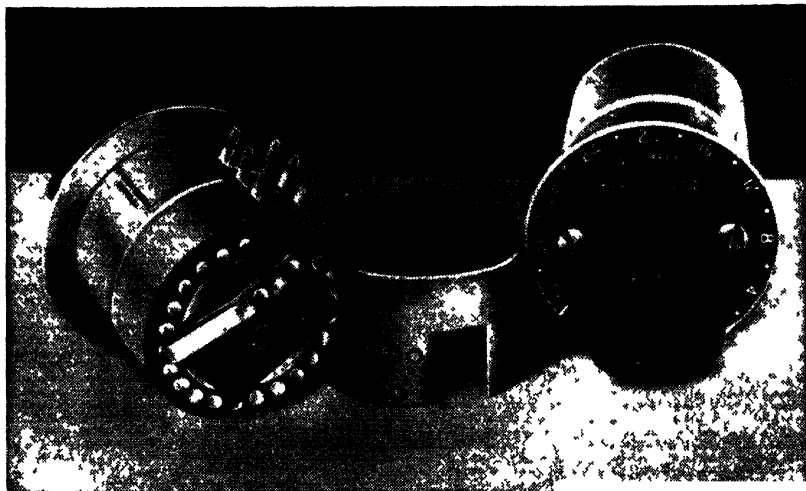
As shown in Fig. 86, a pad is similar in structure to the equivalent circuit for a line of Fig. 67. Thus a pad has a characteristic impedance just as a line has a characteristic impedance. But with a pad, the elements composing it are concentrated in lumps and not distributed as in an open-wire line or a cable. The characteristic impedance of a line (page 130) is the input impedance when the line is terminated in its characteristic impedance. By definition, the **iterative impedance** of a network is the impedance that will terminate a network composed of lumped resistors, inductors, or capacitors in such a way that the input impedance equals the terminating impedance. Strictly speaking, characteristic impedance applies only to lines, in which the resistance, inductance, and capitance are uniformly distributed along the line, but in radio it often is applied to networks, such as pads.

A balanced pad is shown in Fig. 86. The iterative impedance Z_K of such a circuit will be a value of pure resistance equal to

$$R_K = \sqrt{0.25\,R_1^2 + R_1R_2}. \qquad (69)$$

When the pad is terminated in R_K, the ratio of the input voltage E_S to the output voltage E_K is

$$\frac{E_S}{E_K} = \frac{R_K + 0.5\,R_1}{R_K - 0.5\,R_1}. \tag{70}$$

Suppose that it is desired to design a pad to insert a 10-decibel loss in a 600-ohm line. The voltage ratio corresponding to 10 decibels is $E_S/E_K = 10^{0.05 \times 10}$ or $E_S/E_K = 3.162$ (page 96). Using Eq. (70), $3.162 = (600 + 0.5R_1)/(600 - 0.5R_1)$, and $R_1 =$

An attenuator of the type used in radio speech-input equipment. (*The Daven Co.*)

622.5 ohms. The value of R_2 can be found from Eq. (69); thus, $600 = \sqrt{0.25 \times (622.5)^2 + 622.5, R_2}$ and $R_2 = 422$ ohms.

Pads that are variable so that the loss can be adjusted to various settings in decibels are called **attenuators.** They sometimes are arranged to have a value of iterative impedance that approximately is independent of the loss settings. They can be designed step by step, using the same general method as for pads. Attenuators usually are arranged so that all the resistors are variable and are changed by turning one knob.

Filters.—These are networks that are inserted in circuits to suppress certain frequencies and pass others. They are composed of coils and condensers, and since these are lumped rather than distributed, filters have an iterative impedance. This must equal the impedance of the circuit in which the filter is to be used or

reflection will result. For approximate work the coils and condensers used to make filters are assumed to be lossless, and the cutoff frequency (frequency beyond which the filter will not pass) and the characteristic impedance are calculated on this basis.

Low-pass filter sections are shown in Fig. 87. These pass without serious attenuation all frequencies from zero up to the cutoff

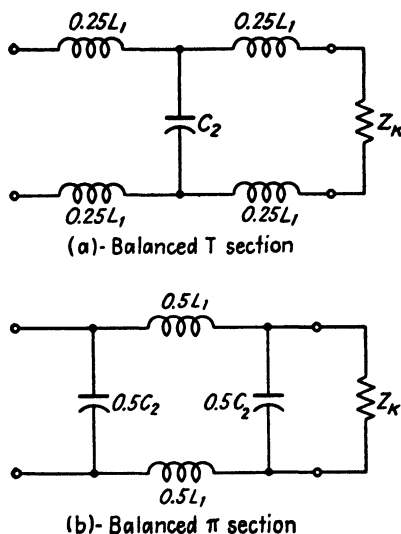



FIG. 87.—Balanced T and π low-pass filter sections. If unbalanced sections are desired, then two coils of $0.5L_1$ henry inductance should be used in one side only of the T section, and one coil of $1.0L_1$ henry inductance should be used in one side only of the π section.

frequency, beyond which the attenuation increases rapidly. The cutoff frequency is given by the equation

$$f_c = \frac{1}{\pi\sqrt{L_1C_2}}. \tag{71}$$

The other equations needed are

$$Z_K = \sqrt{\frac{L_1}{C_2}},\ L_1 = \frac{Z_K}{\pi f_c} \quad \text{and} \quad C_2 = \frac{1}{\pi f_c Z_K}, \tag{71a}$$

where Z_K is in ohms and largely is pure resistance over the band passed, L_1 is in henrys, C_2 is in farads, and f_c is in cycles. Thus a low-pass filter that is to cut off at 1000 cycles and is to be used in 600-ohm circuits would have the values $L_1 = Z_K/(\pi f_c) = 600/(3.14 \times 1000) = 0.191$ henry, and $C_2 = 1/(\pi f_c Z_K) = 1/(3.14 \times$

1000 × 600) = 0.53 × 10^{-6} farad. The actual section then can be made in either of the ways shown in Fig. 87.

High-pass filter sections are shown in Fig. 88. These (theoretically) pass without serious attenuation all frequencies from infinity down to the cutoff frequency, beyond which the attenuation increases rapidly. The cutoff frequency is given by the equation

$$f_c = \frac{1}{4\pi\sqrt{L_2C_1}}. \tag{72}$$

The other equations needed are

$$Z_K = \sqrt{\frac{L_2}{C_1}},\ L_2 = \frac{Z_K}{4\pi f_c} \quad \text{and}$$

$$C_1 = \frac{1}{4\pi f_c Z_K}, \tag{72a}$$

where Z_K is in ohms and largely is pure resistance over the band passed, L_2 is in henrys, C_1 is in farads, and f_c is in cycles. Thus a high-pass filter that is to cut off at 1000 cycles and is to be used in 600-ohm circuits would have the values in Fig. 88 of $L_2 = Z_K/(4\pi f_c) = 600/(4 \times 3.14 \times 1000) = 0.0477$ henry, and $C_2 = 1/(4\pi f_c Z_K) = 1/(4 \times 3.14 \times 1000 \times 600) = 0.132 \times 10^{-6}$ farad.

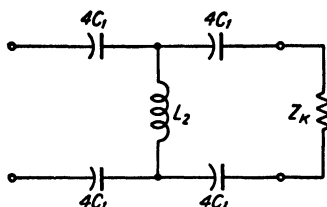
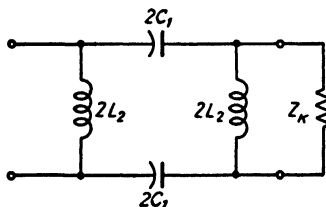



FIG. 88.—Balanced T and π high-pass filter sections. If unbalanced sections are desired, then two condensers of $2C_1$ farad capacitance should be used in one side only of the T section, and one condenser of $1.0C_1$ farad capacitance should be used in one side only of the π section.

Band-pass and Band-elimination Filters are arranged as in Fig. 89. Filters of this general type are used in carrier-telephone systems to separate one channel from another. For this purpose their design is somewhat involved and will receive no further consideration.

The band-pass filter of Fig. 89 will pass a signal frequency to which the series and shunt elements are resonant. At resonance, the series elements offer low impedance, the shunt element offers high impedance, and thus the signal passes through with little attenuation. The usual resonance equations apply for calculating the resonant frequency.

The band-elimination filter of Fig. 89 will pass signals above and below a certain value, but will attenuate the signal frequencies for which the series and shunt elements are in resonance. At resonance the series elements offer very high impedance, and the shunt element offers very low impedance, and thus the signal cannot pass through. The usual resonance equations apply for calculating the resonant frequency.

From the preceding discussion it may appear that only a *single* frequency would be passed, and that only a *single* frequency would be eliminated, but this is not true. In radio, the signal under consideration often is a modulated carrier (page 407), and it consists of the carrier frequency and the two sidebands. In general, it is impossible to make a circuit which uses coils and condensers so "sharp" that it will act only on a single frequency, and most radio circuits of the type of Fig. 89 would pass a band, or reject a band, of about the width of a modulated carrier and its sidebands. Thus these circuits are useful for selecting and rejecting the signals from radio stations. They can be simplified rather than used as shown. Simple resonant circuits used to filter out or reject certain radio stations are sometimes called **wave traps.**

FIG. 89.—The band-pass filter will pass a band of certain frequencies but will not pass other frequencies. The band-elimination filter will not pass a band of certain frequencies but will pass all other frequencies.

At the high frequencies used in carrier-telephone circuits, and at radio frequencies, quartz crystals are used to replace the resonant circuits composed of coils and condensers. This is because at such frequencies the crystal circuits have much lower losses and fewer residual effects than coils and condensers, and thus provide circuits with superior performance. The subject of crystal filters is too specialized to study in detail. The basic idea is illustrated in Fig. 90. The quartz crystal is arranged, as indicated, with four metal electrodes. Often these are plated on the crystal. If the frequency of the incoming signal is varied, the mechanical char-

acteristics of the crystal are such that it vibrates at some frequency, but it will not vibrate much at other frequencies. The internal molecular structure of quartz is such that when it vibrates a voltage will be produced between the output electrodes. Thus at the frequencies at which the crystal vibrates, a signal passes from the input to the output, and the crystal is equivalent to the tuned coupled circuit of Fig. 90. Crystals are used in many other ways, the combination just given being that of a simple band-pass filter.

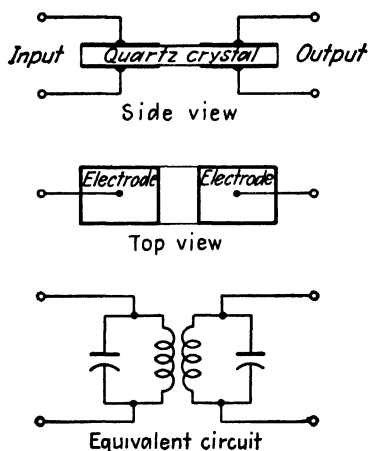



FIG. 90.—Quartz crystals arranged with suitable electrodes are equivalent to sharply tuned inductively coupled circuits.

Equalizers.—Sometimes equalizers are connected between an audio-frequency transmission line, or cable, and the termination in order to make a correction for the fact that the attenuation offered by the line or cable may not be the same at all frequencies over the audio band. Equalizers often are built to correct the distortion on some specific circuit. In practice, an equalizer often is developed through experimentation rather than by exact equations.

As an illustration, suppose that tests on a network channel (telephone circuit) show that the audio-signal volume delivered to the radio transmitter falls off badly at the upper frequencies,

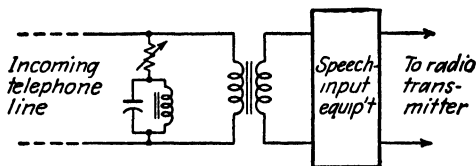



FIG. 91.—A simple series-parallel circuit can be used as an equalizer to improve the high-frequency response of a transmission circuit.

say above 5000 cycles. To remedy this, a coil which has been designed for audio-frequency circuits and which has an inductance of about 0.4 henry is connected in parallel with a paper condenser of about 1.0 microfarad capacitance. This combination will be in parallel resonance at about 8000 cycles. They are connected in

series with a variable wire-wound resistor of several hundred ohms maximum value, and then are connected across the line.

This arrangement is shown in Fig. 91. A voltmeter (copper oxide or vacuum-tube type) is placed across the equalizing network, and an oscillator is connected at the distant sending end. The resistance is varied until the received signal at different frequencies is substantially the same. Some experimentation may be required to find the proper coil, and perhaps it may be necessary to place a resistor in series with one of the parallel branches. The arrangement of Fig. 91 unbalances the circuit slightly, and it may be necessary (to reduce noise and crosstalk) to place the parallel part in the center with half the series resistance on each side of it.

SUMMARY

Audio-frequency transmission lines and cables are used to convey program material and messages to the radio transmitter. Radio-frequency transmission lines and cables are used to connect the radio transmitter to the transmitting antenna. Sections of radio-frequency lines and cables are used to provide inductive or capacitive reactance, and for other purposes.

In a sense, a transmission line or cable merely directs electromagnetic waves to the distant station. If an electromagnetic wave once is created on a line, the wave continues to travel along the line until the energy in the wave is absorbed in line losses, or absorbed in the termination. Some radiation occurs at radio frequencies.

If a transmission circuit is short physically, and the frequency reasonably low, the voltages (and currents) at various points are approximately in phase, and the line is electrically short. If the frequency is increased to a very high value, then the voltages (and currents) are out of phase, and the circuit is electrically long.

Characteristic impedance is the input impedance of a line or cable with distributed (not lumped) constants when the circuit is infinite in length, or terminated in its characteristic impedance. The general equation is

$$Z_O = \sqrt{\frac{R + j\omega L}{G + j\omega C}},$$

where Z_O is the characteristic impedance and R, L, G, and C are the line constants. For radio frequencies, ωL and ωC become so large compared with R and G that the characteristic impedance becomes

$$Z_O = \sqrt{\frac{j\omega L}{j\omega C}} = \sqrt{\frac{L}{C}},$$

and is pure resistance because the $j\omega$ terms cancel. For maximum power transfer and no reflection, a transmission circuit must be terminated in its characteristic impedance.

The line attenuation and the wave velocity are determined by the propagation constant

$$\gamma = \alpha + j\beta = \sqrt{(R + j\omega L)(G + j\omega C)} \quad \text{and} \quad V = 2\pi f/\beta,$$

where γ is the propagation constant and is composed of the attenuation constant α and the phase constant β. The magnitudes of these are determined by the line constants R, L, G, and C. The wave velocity is determined by β, and approaches the velocity of light as a limiting value.

Hard-drawn copper is used for telephone wires and annealed copper for telephone cables. The resistance at audio frequencies is not greatly affected by skin effect. The audio-frequency conductance includes losses caused by the electric field. The cables used for audio purposes often are loaded by adding inductance. This reduces the attenuation and produces other desired results.

Open-wire lines at radio frequencies have greater series resistance R and shunt leakage G than at audio frequencies. The equations for Z_O and γ can be much simplified if certain assumptions are made. For an open-wire line at radio frequencies

$$Z_O = 276 \log_{10} \frac{b}{a},$$

and is pure resistance. The attenuation constant is

$$\alpha = \frac{R}{2Z_O}, \quad \text{where} \quad R = \frac{8.4\sqrt{f}}{a}.$$

The velocity is assumed to be either that of light or perhaps 97 or 98 per cent of that value.

For coaxial cables with air insulation, at radio frequencies the characteristic impedance is

$$Z_O = 138 \log_{10} \frac{b}{a},$$

and is pure resistance. The attenuation constant is

$$\alpha = \frac{R}{2Z_O}, \quad \text{where} \quad R = 4.2\sqrt{f}\left(\frac{1}{a}\right) + \left(\frac{1}{b}\right).$$

Lines and cables usually are terminated in their characteristic impedance when they are used to transmit signal energy between two points. When so terminated, the voltage and current along the line decrease logarithmically. When a line or cable is not terminated in its characteristic impedance, the initial and reflected waves along the circuit interfere and cause regular variations in the resultant voltage and current along the line. These variations are defined as "stationary waves" and usually are called "standing waves" in radio.

The termination of the circuit at the distant end determines the conditions on that circuit at a given frequency. If the circuit is open, a high voltage exists at that point and the current is zero; if the circuit is short-circuited, zero voltage exists and a high current flows through the short circuit.

When open-circuited lines, or cables, are used as circuit elements, the input impedance of a section less than ¼λ long is

$$Z_{OC} = Z_O \cot \frac{2\pi X}{\lambda} \angle -90^\circ,$$

and is capacitive reactance. For a similar short-circuited line

$$Z_{SC} = Z_O \tan \frac{2\pi X}{\lambda} \angle +90^\circ,$$

and is inductive reactance.

Lines or cables that are open- or short-circuited behave like resonant circuits and are called "resonant lines." These lines can be used as impedance-matching transformers.

When lines or cables are not open- or short-circuited, but are terminated in impedances different from Z_O, partial reflection results and the so-called "standing waves" are not complete. The relation is

$$\text{Standing-wave ratio} = \frac{\text{maximum voltage}}{\text{minimum voltage}} = \frac{Z_O}{Z_R} \quad \text{or} \quad \frac{Z_R}{Z_O}.$$

The proper ratio to be used can be determined from a study of the voltage distribution. This equation can be used to measure an unknown impedance.

Pads are used to introduce a known loss into a circuit. The iterative impedance of a network having lumped R, L, C, and G is a term which corresponds to the characteristic impedance of a circuit, such as a line or cable, which has distributed constants.

Filters are inserted in circuits to pass certain frequencies and reject others. These also have an iterative impedance which must match the line or cable in which a filter is inserted. The basic types are the low-pass, the high-pass, the band-pass, and the band-elimination filters. Filters that use quartz-crystal elements are employed.

Equalizers are used to compensate for lack of uniformity in the frequency-attenuation characteristics of a line or cable.

REVIEW QUESTIONS

1. In what ways are open-wire lines and cables used in radio?

2. Explain why two short parallel rods may be classed as a radio-frequency transmission line.

3. Explain why a study of voltage and current on a transmission line may be used to determine electromagnetic wave phenomena.

4. Give two definitions for characteristic impedance.

5. Why should a circuit be terminated in its characteristic impedance?

6. Explain the difference between a balanced and an unbalanced transmission circuit, and give an example of each.

7. How does the transmission of audio-frequency signals over open-wire lines compare with transmission over cables?

8. Why does the series resistance R of radio-frequency circuits differ at radio frequencies from the audio-frequency value?

9. What is the nature of the characteristic impedance of a radio-frequency transmission line?

10. What are the features of a coaxial cable that make it such a desirable radio-frequency transmission circuit?

11. Why are many coaxial cables made with a particular ratio of b/a?

12. Are there instances in which wave reflection on radio-frequency lines or cables is useful? If so, name several.

13. A lossless line is three-fourth wavelength long and open at the distant end. What are its characteristics?

14. Repeat Question 13 for a shorted line.

15. Discuss the nature of the input impedance of a lossless line that is open or shorted at the distant end.

16. Explain how a section of a line or cable could be used as a coil or condenser.

17. What is meant by the term "resonant line"?

18. Explain why a resonant line may be used as a transformer.

19. Why is it possible to use a quarter-wave metal bracket as an insulator?

20. At what other frequencies could the metal bracket be used as an insulator?

21. If a lossless line contains a resistance component in its termination, what must be the nature of the input impedance? Why?

22. How can the presence of standing waves on a line be detected?

23. Why are pads sometimes used?

24. What is meant by iterative impedance, and to what types of circuits does it apply?

25. What is a crystal filter, and why are crystals used?

PROBLEMS

1. Calculate the resistance per mile at 1000 cycles and 20°C. for a hard-drawn copper open-wire line composed of two wires 0.104 inch in diameter.

2. Calculate the inductance per mile of the line of Prob. 1 if the spacing is 12 inches.

3. Calculate the capacitance per mile of the line of Probs. 1 and 2.

4. Using the values previously determined, calculate the characteristic impedance, the attenuation in decibels, and the wave velocity.

5. If the line previously considered is 25 miles long, if 1.0 volt at 1000 cycles is impressed at the sending end, and if the distant end is terminated in a resistor equal to the characteristic impedance, calculate the approximate input current, input power, output current, output voltage, and output power.

6. An open-wire radio-frequency transmission line is 200 feet long and is composed of two No. 10 A.W.G. hard-drawn copper wires held 6.0 inches apart by ceramic spreaders spaced every 18 inches. The system is to operate at 12.0 megacycles. Calculate the characteristic impedance and the total attenuation in decibels. The power input to the line is 250 watts. Calculate the power lost and the efficiency of the line.

7. Repeat the calculations made in the illustrative problem on page 140 at a frequency of 30 megacycles.

8. Repeat the calculations made in the illustrative problem on page 145 at a frequency of 65 megacycles.

9. Repeat the calculations made on page 157 if the transmission line has a characteristic impedance of 600 ohms.

10. Calculate the values of R_1 and R_2 for a pad to introduce a 15-decibel loss into a 600-ohm circuit.

11. Determine the values of L_1 and C_2 for a low-pass filter to work in a 600-ohm line and cut off at 10,000 cycles. Show the values for both T and π sections.

12. Repeat Prob. 11 for a high-pass filter.

CHAPTER VI

VACUUM TUBES

The extent to which vacuum tubes are used in radio is well known. Modern radio systems could not exist in their present form without vacuum tubes.

These tubes may be regarded in two ways: (*a*) from the electronic viewpoint, considering what happens within the tubes; and (*b*) how the tubes function in circuits, considering their performance as circuit elements. In this chapter vacuum tubes will be considered from the electronic viewpoint. They will be considered as circuit elements in the chapters that follow.

The term "vacuum tube" is used to include both **high-vacuum tubes** and **gas tubes.** High-vacuum tubes are evacuated to the extent that the residual gas will not appreciably affect their operation, and are then sealed. Gas tubes also are evacuated, but before they are sealed a small amount of some stable gas (such as argon) is admitted. In many tubes, particularly those used in rectifiers, a small amount of mercury in liquid form (which vaporizes) is placed in the tube. These also are classed as vacuum tubes.

In vacuum tubes, photocells, and other electronic devices it is necessary to free electrons from the metal electrodes within the tube. These **electrons** are small negative particles of electricity, and they exist in metals in vast numbers. These negative electrons may be freed from the electrodes by the application of energy to the electrodes. In the devices used in radio, this is accomplished in several ways as follows:

a. Thermionic Emission.—When heat energy is applied to a metal and its temperature is raised to a sufficient point, electrons are given off by the metal.

b. Secondary Emission.—A rapidly moving beam of electrons contains sufficient energy to liberate **secondary electrons** from a metal when such a beam is directed against a metallic surface.

c. Field Emission.—If a cold metallic surface is subjected to a *very* strong electric field, electrons are liberated from the metal.

d. Photoelectric Emission.—The energy in a beam of light can release electrons from a metal if a suitable beam is directed against a metallic surface.

Thermionic Emission.—This method is used to liberate electrons in the vacuum tubes used in radio for such purposes as rectification and amplification. Such tubes contain a **thermionic cathode** which is heated by an electric current to a temperature such that this cathode emits negative electrons into the region surrounding the cathode. These electrons then are acted upon by the other electrodes (such as the grid and plate) within the tube, and the tube is made to rectify, amplify, and perform many useful functions.

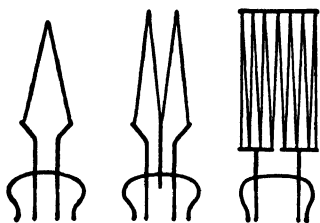



FIG. 92.—Typical filament-type cathodes used in small vacuum tubes.

The emission of negative electrons from a hot cathode has been likened to the evaporation of a liquid by the application of heat energy. Of course there is a difference, because in vacuum tubes the electrons are returned by the external circuit to the cathode. Thus a thermionic cathode does not "dry up," electronically speaking, as does heated liquid in an open vessel.

The thermionic cathodes used in vacuum tubes are of two general types: (*a*) the **filament type** in which the heating current passes directly through the metallic wire, or ribbon, constituting the cathode, and this filament emits the electrons; and (*b*) the **separate-heater, or indirectly heated type,** which often is a metallic cylinder through which a heater wire passes. The cylinder usually is coated with a substance that readily emits electrons.

Filament-type Cathodes.—Typical filament-type cathodes are shown in Fig. 92. The electric heating current is supplied by an external source and passes through the wires. These wires, or ribbons, sometimes are of pure metal, such as tungsten, sometimes are of thoriated tungsten, and sometimes are a pure metal wire or an alloy, with an electron-emitting oxide coating applied to the surface.

Pure Metal Filaments.—Tungsten is so universally used for this purpose that other metals need not be considered. Tungsten filaments are used in the high-vacuum high-power transmitting tubes for the final stages of certain large radio transmitters. Tubes with tungsten filaments are not used in the common types of radio

receivers. One of the important reasons tungsten is used in large tubes is that such tubes operate at high voltages, and the tubes must be evacuated to a very high degree or the residual gas may affect the operation of these tubes. Tungsten is very stable and does not, during the evacuating process or in later operation, tend to throw off gas particles from itself; thus these tubes may be highly evacuated, and will retain their high degree of vacuum. A second reason is that in power-transmitting tubes the high voltages tend to ionize (page 201) some of the remaining gas (all of it cannot ever be pumped out), and this ionization process produces both positive and negative ions. The positive ions are quite massive, and flow to the negative filament. These ions gain much energy, and when they reach the filament they bombard it. Tungsten can stand such bombardment without disintegrating. A third reason that tungsten is used is that it has a very high melting point, and thus it can be operated without damage at the high temperatures necessary to supply the required electron emission. The operating temperature is about 2500°K. (degrees Kelvin).[1] From the standpoint of the heating energy required, pure tungsten is not a very efficient electron emitter when compared with other materials. But again, tungsten will stand up well in high-voltage high-power transmitting tubes, whereas most other materials will not.

Thoriated-tungsten Filaments.—During the manufacturing process, a small amount of thorium oxide is added to the tungsten used for the wire in these filaments. After the filament is in place, and during the manufacturing process, the thoriated-tungsten filament is heated to a high temperature. This "boils" *some* of the thorium to the surface where it forms a *thin* layer on the filament. This thin layer of thorium assists the electrons in escaping from the metal below. The exact action is somewhat obscure, but whatever its nature, the effect is very pronounced. It is thought possible that the thin thorium layer acquires a positive charge that assists the negative electrons in escaping from below. The energy that electrons must have to escape from a metal is called the **surface work function.** The thin layer of thorium, in effect, lowers this surface work function. The layer must be thin so that the electrons from below can pass out readily between the thorium atoms. Under comparable conditions the emission from a thoriated-

[1] Degrees centigrade plus 273.

tungsten filament is perhaps one thousand times greater than for pure tungsten. Thoriated-tungsten filaments have been improved greatly in recent years by a **carburizing process** that hardens the surface. Such filaments now are used extensively in large power-transmitting tubes, and in a sense compete with pure tungsten filaments. Among other factors, better evacuation makes this possible. Thoriated-tungsten filaments are operated at about 1500°K.

Oxide-coated Filaments.—These consist of a wire, or ribbon, of nickel or some suitable alloy on which a coating of barium and strontium is placed. After the filament is mounted, and during manufacture, the filament is activated by heating it to a high temperature. This "boils" a thin layer of barium to the surface, and this layer gives the filament extraordinary emitting characteristics. The strontium gives the coating mechanical strength. The thin layer of barium atoms appears to acquire a positive charge that lowers the surface work function and aids the electrons to escape from beneath. This increases the emission millions of times over that of the pure metal. Oxide-coated filaments are used in small tubes, such as those for radio receivers and very small transmitters. Filaments of this type are used also in gas and mercury-vapor rectifiers and similar tubes. Oxide-coated filaments are not used in medium- or large-sized high-voltage high-power transmitting tubes, because it is difficult to obtain the extremely high vacuum needed, and because the oxide coating is not so rugged as the other surfaces. Oxide-coated filaments are operated at about 1000°K.

Fig. 93.—Separate-heater type cathodes used in small vacuum tubes.

Separate-heater Cathodes.—When alternating current is used to heat tubes with filament cathodes, hum is likely to be produced. This is particularly true in the first tubes of a radio receiver, where the received signal strength is low. Also, in circuit design it often is advantageous to have a heater that is electrically insulated from the emitting cathode. For these and other reasons many tubes, particularly the small ones, use a cathode with a separate heater.

Two types of separate-heater cathodes are shown in Fig. 93; other variations are possible and are used particularly in tubes for industrial rather than for radio purposes. As this figure indicates,

these cathodes are made of metal tubes with insulated heater wires of pure tungsten at the center. Oxide coatings are placed on the surfaces of the metal tubes. The heater wires are often spiraled, as indicated, and are insulated and suspended so that when in place within the oxide-coated cylinder there is no electric contact between the heater and the tube.

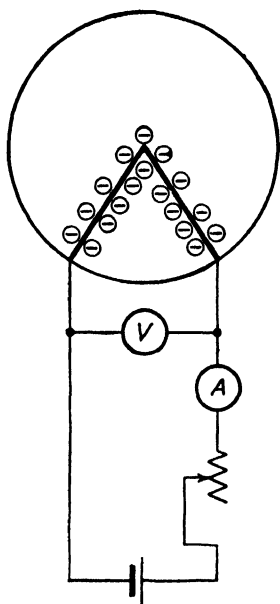



FIG. 94.—Negative electrons are given off by the heated filament in a vacuum tube. These electrons form a random negative space charge about the filament.

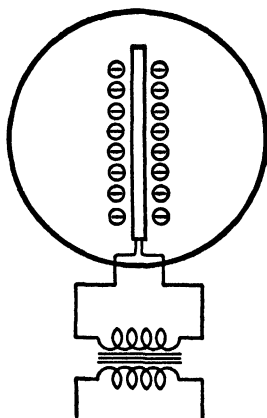



FIG. 95.—A filament transformer often is used as shown to heat a cathode. In this illustration the cathode is of the separate-heater type. The emitted electrons form a negative space charge about the cathode. Actually, the individual electrons occupy positions at random, instead of being aligned as shown in Figs. 94 and 95.

The Negative Space Charge.—When the filament (or cathode) of a tube is heated, electrons are emitted into the region immediately surrounding the filament. This is shown in an elementary way by Fig. 94. The ammeter indicates the current flow and the voltmeter measures the voltage drop; from these, the power consumed by the filament can be computed. The emitted electrons surrounding a separate-heater cathode are shown in Fig. 95. The heating current is supplied by a small filament-heating transformer usually called a **filament transformer** that steps down from 110

volts to the 6.3 volts (or other value of alternating voltage required) to heat the cathode.

The tubes shown contain only a cathode. When the negative electrons are thrown off by the heated cathode, they take negative electricity away from the cathode, thus leaving the cathode positively charged. The positively charged cathode immediately attracts the negative electrons, and tends to pull them back. Thus in a tube containing only a cathode, an equilibrium condition soon is established in which the number of electrons thrown off by the hot cathode equals the number of electrons pulled back by the cathode. At any instant there are, however, many electrons in the space around the cathode. This accumulation of negative electrons in the space around a hot cathode is called the **negative space charge,** or **space charge.** (As will be seen later, positive space charges may exist in gas tubes.)

FIG. 96.—The diode consists of a plate and a cathode. The cathode may be a filament or may be indirectly heated as this figure shows. The space charge is not shown on circuit drawings. It is included here to emphasize its presence in a tube. The transformer steps down the voltage from, say, 110 volts to 6.3 volts for heating the cathode.

Actually, the negative space charge extends throughout the tube, but the intensity is so much greater close to the cathode that it is convenient to consider that the space charge is confined to a thin sheath immediately surrounding the cathode.

The Diode.—This is a **two-electrode tube** containing a cathode and an anode, or plate. The conventional way of drawing a diode is shown in Fig. 96; actually, the plate usually surrounds the cathode in high-vacuum diodes. In some gas diodes, however, the plate may be suspended above the cathode, as Fig. 96 indicates. The space charge is not shown on circuit diagrams. The discussion that follows applies to high-vacuum diodes.

If the plate is made positive by the battery and voltage divider of Fig. 96, negative electrons will be drawn over to the positive plate from the negative space charge. When they leave the space-charge region, the condition of equilibrium mentioned in the preceding section is upset, and new electrons from the hot cathode enter the space-charge region. A new condition of equilibrium is reached then, in which the number of electrons entering the space-charge region equals the sum of those flowing back to the cathode

plus those flowing to the positive plate. Those electrons which are attracted to the positive plate cannot accumulate on the plate, and therefore pass through the milliammeter and on back to the cathode. A flow of negative electrons down through the milliammeter is equivalent to a flow of conventional electric current up through the milliammeter. It must be remembered that the conventional direction of current flow is from positive to negative, and that this was set as standard before the science of electronics was developed. Thus *current flow* always is opposite to *electron current flow*.

In studying a diode, it always will be assumed that the cathode is at normal operating temperature, as is true in practice. The characteristics of most importance then become the way in which the plate current (as read by the milliammeter of Fig. 96) varies as the plate voltage is changed by adjusting the voltage divider. These variations are shown in Fig. 97.

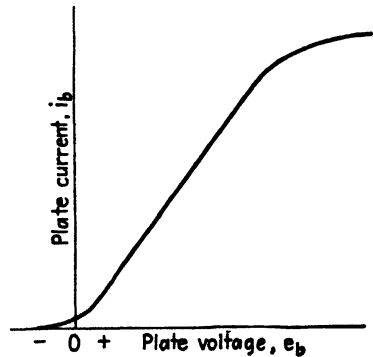



FIG. 97.—Variations of plate current with plate voltage for a high-vacuum diode.

When the plate is at zero potential, or even slightly negative, some electrons flow to it. The reason for this is that some of the electrons, when thrown off the cathode, have initial velocities sufficient to carry them over to the plate, even if the plate is at zero potential, or slightly negative *with respect to the cathode.* In studying vacuum tubes the negative cathode terminal is taken as the reference point for measuring the electrode voltages if batteries are used to heat the filaments. If a transformer is used to heat the filament-type cathode, then the center tap on the secondary of the filament transformer is the reference point. If the tube has a separately heated cathode, then the cathode (metal tube on which the oxide is placed) is the reference point.

As the plate of the diode of Fig. 96 is made more positive with respect to the cathode, more electrons are attracted over to the plate. As indicated by Fig. 97, the plate current will increase. This cannot increase indefinitely, however, and a voltage is reached at which the plate current increases but little, if at all. This is because the plate is taking the electrons as fast as the hot cathode

is emitting them. This fact causes the plate-current curve of Fig. 97 to flatten off, and the condition often is referred to as **voltage saturation.** In other words, for the conditions applying, the tube has all the voltage on the plate that "it can handle." Increasing the cathode temperature would increase the plate current, but, as previously mentioned, these discussions assume that the cathode is at normal operating temperature.

The Triode.—As explained in the preceding section, the hot cathode emits negative electrons, and these form in a sheath, or space charge, about the cathode. The positive plate pulls negative electrons from the space-charge region, and these electrons flow on around the electric circuit and back to the cathode. This constitutes the **plate current.** The following discussion applies to *high-vacuum* triodes.

In the **triode,** or **three-electrode tube,** a **grid** is inserted between the plate and cathode. The grid sometimes is a wire mesh but more often is a spiral of wire (Fig. 106, page 193). The grid is closer to the supply of electrons than is the plate, and furthermore, the electrons flowing to the plate must all pass through the grid. For these reasons the grid is able to control, to a great extent, the flow of electrons to the plate. In this way the grid also controls in a large measure the plate current in the external circuit.

The circuit of Fig. 98 is useful in studying the control effect exerted by the grid. Assume that the cathode is heated to normal operating temperature, and that the plate is at the normal positive value. The supply circuit for the heater is not shown; this is standard practice, it being assumed that the heater is properly connected and energized. If the reversing switch is "up" (as shown by the dotted lines), and the voltage divider across the *C* battery is set as indicated, the grid will be made *positive* with respect to the cathode. The positive grid will neutralize some of the negative space charge and will attract electrons toward itself. Some of these electrons will strike the positive grid wires, and will flow through the connected grid circuit and back to the cathode. Many of the electrons now in rapid motion will miss the grid wires and will flow on to the positive plate and on around the plate circuit and back to the cathode. In this way, the grid *increases* the flow of plate current.

If the switch (Fig. 98) is thrown "down," the grid is made negative with respect to the cathode. The grid now will repel electrons

and in a sense assist the negative space charge in forcing electrons *back toward the cathode.* In this way the negative grid opposes the action of the positive plate, and reduces the current that otherwise would flow to the plate. The grid is the *control electrode* in the triode.

The action just explained can be represented by the set of curves of Fig. 99. Assume that the cathode (Fig. 98) is at normal operating temperature and that the plate is at normal positive operating

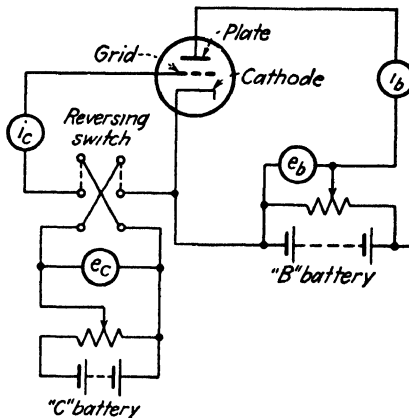



FIG. 98.—Circuit for studying the control effect of the grid in a three-electrode high-vacuum tube or triode.

voltage e_b. If the grid is made sufficiently positive, grid current will flow as indicated by e_b of the lower set of curves, and plate current will flow as indicated by e_b of the upper set of curves. As the grid is made less negative, the currents change as shown by these curves. When the grid is at zero potential (the same potential as the cathode), the grid current falls almost to zero, but a very small amount may flow because of the velocity of the emitted electrons (page 179). At zero grid voltage the plate current curves cross the zero axis as indicated. If now these tests are repeated but with a *higher* plate voltage, then the plate will attract more of the electrons and curves marked e_b' result. Note that the grid current is less than before, because the plate with its increased potential is now more effective in pulling the electrons through the grid and less electrons strike the grid. If now these tests are repeated with the plate at a *lower* voltage than in the first tests, the resulting currents will be as shown by e_b''. Note that the grid

current now is greater than in the first test because the plate with its decreased potential is less effective in pulling electrons through the grid and more electrons strike the grid.

The peculiar shapes of the grid-current curves of Fig. 99 is caused by *secondary emission* (page 173). For a small radio tube with normal positive plate voltage of perhaps 90 volts, the grid, when it is about 10 volts positive, draws an appreciable current, and the electrons constituting the current flow achieve considerable velocities. Because of this they have energies sufficiently high to knock out secondary electrons when they strike the wires of the grid. Some of these secondary negative electrons will be attracted back to the positive grid, but many of the secondary electrons will flow over to the more positive plate. Thus electrons are flowing *to* the grid from the cathode, and *away* from the grid to the plate. This loss of electrons causes the grid current to drop. If, however, the grid is made more positive, the net grid current rises because the grid is able to capture and "reclaim" most of the secondary electrons.

FIG. 99.—Characteristic curves for a high-vacuum triode. Relations between grid voltage and plate and grid currents at various plate voltages. Plate-current values and grid-current values are not to the same scale.

The Plate Current.—The plate current that flows in a triode under various operating conditions was discussed in the preceding section. As Fig. 99 shows, with the grid positive the plate current rises to high values, and if the grid is sufficiently negative, the plate current falls to zero. These represent the extremes of operation.

Vacuum tubes are used extensively as amplifiers, and for this purpose they often are operated with the grid negative and the plate positive, which confines the operation to the left portion of Fig. 99. This part of the figure has been enlarged in Fig. 100.

If the direct voltage between the grid and cathode is $-E_c$ and the direct voltage between the plate and cathode is $+E_b$, then the resulting plate current will be I_b, as indicated. When a tube is used as an amplifier, two voltages simultaneously are impressed in series between the grid and cathode: (*a*) the direct **grid-bias voltage**

$-E_c$, which fixes the point of operation on the plate-voltage curve, and (*b*) the alternating-current signal voltage E_g. These also are shown in Fig. 100.

The positive half cycle of the signal voltage impressed on the grid subtracts from the negative grid-bias voltage E_c, furnished by the direct-voltage source E_{cc}. The negative half cycle of the signal

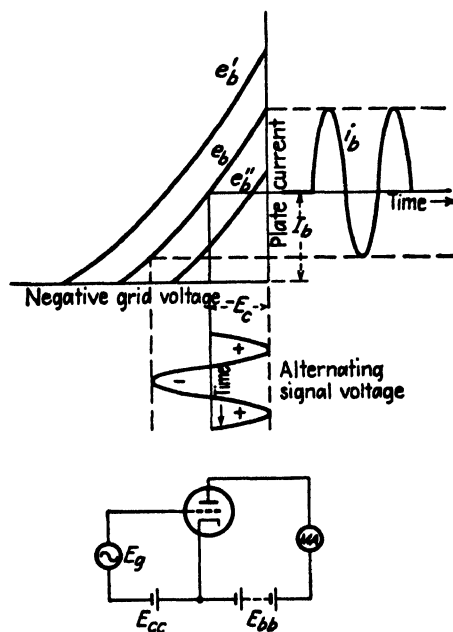



FIG. 100.—If the grid-bias voltage is $-E_c$ and the plate voltage is $+E_b$, then a plate current I_b will be measured by the milliammeter in the plate circuit. If an alternating signal voltage E_g is impressed, the plate current will vary as shown by i_b. This contains a direct and an alternating component. In an amplifier a load of some type (such as a resistor) is placed in the plate circuit.

voltage impressed on the grid adds to the negative grid-bias voltage E_c. As a result of this action, the alternating signal voltage applied to the grid causes the net grid potential to be alternately less negative and more negative. This alternating potential of the grid allows more current and less current to flow, and as a result the plate current will be as indicated by i_b.

It is very important to note that the grid voltage has two components: (*a*) the direct grid-bias voltage that fixes the point of operation at the proper point on the plate-voltage curve deter-

mined by the specific value of voltage source E_{bb}, and (*b*) the alternating signal voltage applied to the grid. Likewise, the plate current contains two components: (*a*) the direct-current component I_b that would be measured by a direct-current milliammeter in the plate circuit, and (*b*) an alternating current that flows because of the grid signal-voltage variations.

The direct grid-bias voltage and the direct plate voltage are of much importance. They must be properly selected or the tube will not operate as desired. However, in much of the design work it is assumed that the direct-voltage values are correct, and attention is concentrated on the alternating components. The alternating components of grid voltage and plate current are the signal being passed through the tube and its associated external circuit.

Vacuum-tube Coefficients.—Vacuum tubes are used in circuits to perform such functions as amplification, etc. If the performance of such circuits is to be predetermined by calculations, then certain tube coefficients or "constants" must be known. These coefficients will be different at different grid and plate voltages. For many purposes, triodes are operated with the grid negative and the plate positive with respect to the cathode. Under these conditions the tube operates over the part of the curves of Fig. 99 to the left of zero grid voltage. This portion of the curves of a typical triode was shown in Fig. 100, and it is shown with numerical values in Fig. 101. These two sets of curves are the same data plotted in different ways. Figure 101*a* is a family of **constant plate-voltage curves,** and Fig. 101*b* is a family of **constant grid-voltage curves.**

Amplification Factor.—For the first set of curves it will be noted that when the plate is at + 90 volts, a grid voltage of about −11 volts reduces the plate current to zero, but when the plate voltage is at a higher value of +120 volts, a grid voltage of over −15 is required to reduce the plate current to the zero value, or **cutoff** as it is called. For the second set of curves it will be noted that when the grid is at −4 volts, a plate voltage of +30 volts must be applied before current flows; also that when the grid voltage is −8 volts, the plate voltage must be +60 volts before plate current flows. Because of its strategic position a few volts on the grid is more effective than many volts on the plate.

The **amplification factor** is a measure of the effectiveness of the grid with respect to the plate. From the preceding values, and for the tube whose characteristics are given by Fig. 101, the grid

is about eight times as effective as the plate in controlling plate-current flow; that is, $\mu = 120/15 = 8$. The Greek letter μ (mu) is used to designate the amplification factor. The value of the amplification factor of a tube depends to some extent on the

(a)-Grid voltage-plate current curves at several values of plate voltage

(b)-Plate voltage-plate current curves at several values of grid voltage

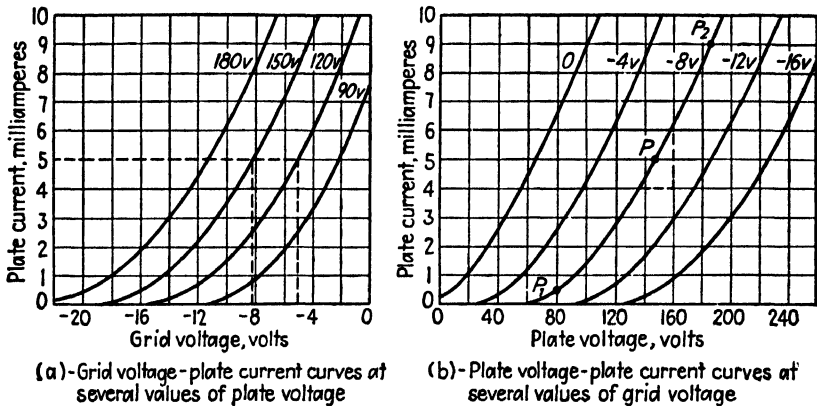



FIG. 101.—Characteristic curves for a triode. The amplification factor is a measure of the relative effectiveness of the grid with respect to the plate. In (a) if the plate voltage is increased from 120 to 150 volts, and if at the same time the grid voltage is decreased from −5 to −8.3 volts, the plate current remains constant at 5 milliamperes. Thus 3.3 volts change in grid voltage is as effective as 30 volts change in plate voltage, and the amplification factor is $\mu = 30/3.3 = 9.1$.

The alternating-current plate resistance, which is the opposition the tube offers to the flow of alternating current, is found from (b). Assume that the tube is being operated at point P with −8 volts on the grid, 147 volts on the plate, and with a plate current of 5 milliamperes. An alternating current would vary about point P, going above P and below P. The length of the base of the triangle drawn at point P is 25 volts, and the height of the triangle is 2.2 milliamperes or 0.0022 ampere. The alternating-current plate resistance is the reciprocal of the slope of the triangle, or $r_p = 25/0.0022 = 11{,}350$ ohms. The plate resistance depends on the point of operation, whether at P, P_1, or P_2. The direct-current plate resistance, as distinguished from the alternating-current plate resistance, at any point P is merely the value of plate voltage divided by the corresponding value of plate current.

The grid-plate transconductance can be found from (a) by erecting a small triangle at the operating point as in (b). Or, the grid-plate transconductance can be found from the relation $g_m = \mu/r_p$. For this tube $g_m = 9.1/11{,}350 = 0.0008$ mho = 800 micromhos.

potentials on the electrodes. The value of the amplification factor μ, as just determined, was from data in the vicinity of current cutoff. Although tubes sometimes are operated near cutoff, more often they are operated along the essentially straight portion of the plate-current curves as shown in Fig. 100. The method of determining the amplification factor in this region is explained under Fig. 101. The amplification factor has no unit of measure.

Plate Resistance.—As explained in the preceding section, when an alternating signal voltage is impressed on the grid of a properly biased tube, the plate current contains both direct and alternating components. For the purpose of study, it may be assumed that these two components exist separately in a tube. Because of the space charge within the tube there is opposition to current flow within the tube. If the direct voltage between the plate and cathode is divided by the direct-current component flowing from cathode to plate (this is the direction of electron flow), a value of resistance is obtained, and this is the direct-current plate resistance of the tube. This value sometimes is of interest, but the alternating-current plate resistance is of most concern (for reasons that will be made clear in the chapters that follow). Plate resistance is measured in ohms.

The **alternating-current plate resistance,** or **plate resistance** as it usually is called, is the internal opposition offered between the cathode and plate to the flow of the alternating-current component. This is sometimes referred to as **plate impedance,** but for all except the high radio frequencies this is essentially resistance with negligible reactance. The plate resistance, designated by r_p and measured in ohms, is found as explained under Fig. 101. It is numerically the reciprocal of the slope of the plate voltage-plate current curve.

Mutual Conductance.—As explained on page 183, when an alternating signal voltage is impressed on the grid circuit, an alternating signal current flows in the plate circuit. Note in particular that the *cause* is in one circuit and the *result* is in a separate circuit. Thus a transfer has taken place. In general, current = voltage × conductance, but when the voltage is in the grid circuit and the current in the plate circuit, the relation becomes current = voltage × transconductance.

The **grid-plate transconductance,** or **mutual conductance,** is a factor by which to multiply the alternating grid signal voltage to obtain the alternating plate current. This factor is represented by g_m and is measured in mhos. It is numerically the reciprocal of the slope of the grid voltage-plate current curve and is evaluated as explained under Fig. 101.

The coefficients defined in this section are *for the tube alone,* and therefore do not take into account the external circuit. Of course in actual circuits some circuit element (such as a resistor) or some

device (such as a loudspeaker) would be connected in the plate circuit. The effects of such elements and devices will be considered in the following chapters.

The tube coefficients or "constants" are not exactly constant, but vary with the direct electrode voltages (direct voltages between plate and cathode, and between grid and cathode). The variations for a typical triode are as shown in Fig. 102. The coefficients of

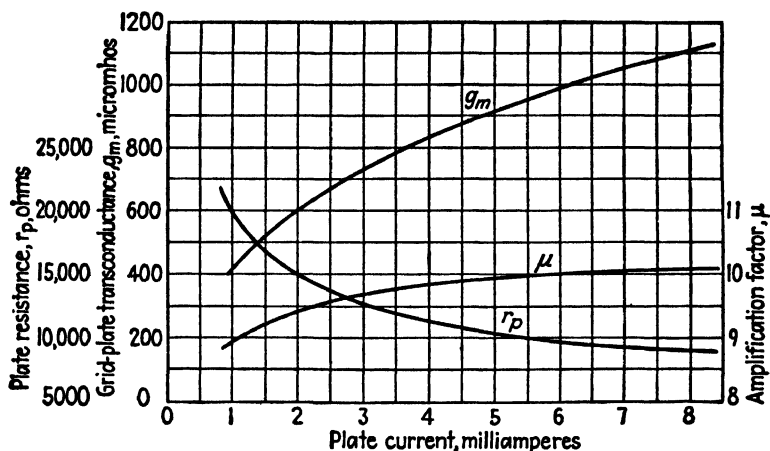



FIG. 102.—Variations in grid-plate transconductance, amplification factor, and plate resistance with variations in plate current. Grid voltage held constant at −8 volts, and plate current varied. The curves are for a typical triode similar to the one represented by Fig. 101.

this figure are for the same tube, may be determined from the same sets of curves, and are related as follows:

$$\mu = r_p g_m, \qquad g_m = \frac{\mu}{r_p}, \qquad \text{and} \qquad r_p = \frac{\mu}{g_m}. \tag{73}$$

A clue for remembering these is as follows: The amplification factor μ has no unit of measure; plate resistance r_p is in ohms, and mutual conductance g_m is in mhos; multiplying ohms × mhos would cancel the units because they are reciprocals; thus, $\mu = r_p g_m$.

The Tetrode.—For many years triodes were used almost exclusively in radio. Today triodes are used extensively in radio transmitters for the final power-output tubes. Modern radio receivers, however, first were made possible by the development of the **tetrode, or four-electrode tube.** Triodes tend to oscillate because of feedback through the interelectrode capacitance between the plate and the grid. Early radio receivers required

special circuits to neutralize this feedback and prevent oscillations. In the tetrode, or four-electrode tube, a **screen grid** is inserted between the **control grid** and plate as indicated in Fig. 103. This shields the control grid from the effect of the plate, and oscillations are prevented to a large extent, except at very high radio frequencies. Note the use of the term "control" grid. This designation is necessary in **multigrid tubes.** The control grid is the grid on which the input signal is impressed.

In studying the triode, it was explained that the grid is much more effective than the plate in controlling the electron-current flow. This is because the grid is between the plate and the space-charge region, and thus the grid shields the plate and renders it somewhat ineffective in pulling electrons from the space-charge region. In a sense, the plate gets what electrons the grid will let it have.

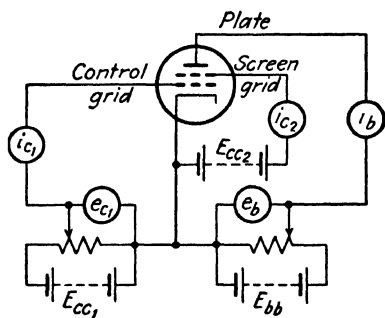



Fig. 103.—The tetrode contains a screen grid between the control grid and plate. The screen grid usually is held at a constant positive potential such as + 90 volts.

Now in a tetrode this effect is increased by the presence of the screen grid. In this tube the plate is shielded by the screen grid and by the control grid, and the plate thereby is rendered even more ineffective in influencing the electron-current flow. The control grid is about as effective as before, and thus the tube has a very high amplification factor, perhaps as high as 500.

In Fig. 101*b*, the reciprocal of the slope of the curves gives the plate resistance. Corresponding curves for a tetrode will be very flat (over the operating range) because the plate, being shielded by the two grids, has little effect on the plate current. Thus the plate resistance of a tetrode is very high, being 500,000 ohms for a typical tube.

Because the control grid in a tetrode has substantially the same effect on the plate current as it does in a triode, the mutual conductance is about the same value as for a triode.

The characteristic curves for a typical tetrode are as shown in Fig. 104. The curve of Fig. 104*a* is similar to the curves of Fig. 101*a*. This is because, as previously stated, the grid of the triode

and the control grid of a tetrode are about equally effective. In taking data for the characteristic curves of a tetrode, the screen grid is held at a constant positive voltage of about +90 volts as would be done in practice. The screen grid therefore acts as an electron accelerator. In fact, it pulls electrons over from the space-charge region and through the negative control grid. Some of these electrons strike the positive screen-grid wires, but others, because of their high velocities, pass between the screen-grid wires and are drawn over to the positive plate.

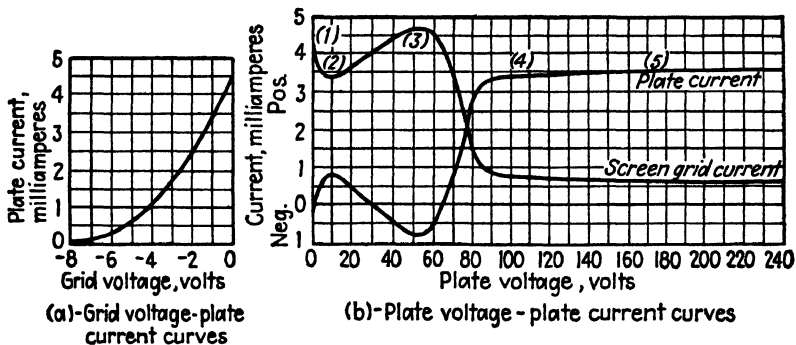



FIG. 104.—Characteristics of a tetrode. The plate is shielded so completely by the screen grid and the control grid that variations in plate voltage have little effect on the space charge and the plate current. For this reason the graphical methods of Fig. 101 cannot be used with accuracy to find μ, g_m, and r_p. Over the region from about 10 to 50 volts the tube exhibits a negative resistance characteristic, an increase in plate voltage causing a decrease in plate current.

The characteristic curves of Fig. 104*b* will be analyzed now, it being assumed that the cathode is at operating temperature, that the control grid is held constant at a normal operating voltage of, say, -3 volts, and that the screen grid is held constant at a voltage of $+90$ volts. A circuit such as Fig. 103 is used to determine the characteristics.

At point 1, the plate is at zero volts and will attract no electrons, but the screen grid is at $+90$ volts and will take a large current. At point 2, the plate is at $+10$ volts and will draw a few electrons that otherwise would go to the screen grid. Since the control grid largely controls the number of electrons that leaves the space-charge region, the sum of the screen-grid current and plate current essentially is constant. Thus at point 2 the screen-grid current falls.

Secondary emission (pages 173 and 182) starts to occur *at the plate* when the plate voltage is at about +10 volts. The rapidly moving electrons striking the plate knock secondary electrons out of the plate. These secondary electrons have two choices: (*a*) They may return to the plate which is at +10 volts, or (*b*) they may travel over to the screen grid which is at +90 volts. Naturally, many of the secondary electrons go to the more highly positive screen grid.

As the plate is made more positive, say +20 volts, the electrons are further accelerated, more secondary emission occurs, and the screen grid draws more electrons from the plate. The screen grid now has two sources of electrons: (*a*) from the cathode and (*b*) from the plate. As a result, in the vicinity of point 3 the screen-grid current rises, and since the plate is losing secondary electrons, the plate current falls. These facts can be verified by checking the directions of current flow on Fig. 103. The normal electron current flows to the plate, down through the plate milliammeter, and back to the cathode. The normal electron current flows to the screen grid down through the screen-grid milliammeter and to the cathode. The secondary electrons flow to the screen grid down through the screen-grid milliammeter, up through the plate milliammeter, and back to the plate. Thus the screen-grid milliammeter is affected by two components in the *same* direction and the reading increases, but the plate milliammeter is affected by two components in *opposite* directions and the current decreases.

At point 4, the plate has been made more positive than the screen grid, and the plate is now able to attract all the secondary electrons back to it. Also, the plate is able to draw to itself many electrons that otherwise would go to the screen grid. As a result, at point 4 the plate current has increased and the screen-grid current has decreased.

At point 5, the plate voltage has been made very much larger than the screen-grid voltage. As a result, the action of the preceding paragraph has become more pronounced. The plate-current curve "flattens out" because the plate is not effective in pulling electrons out of the space-charge region. The effect of the plate largely is limited to attracting electrons after they pass through the screen grid.

The curves of Fig. 104 are for an early tetrode in which secondary emission was excessive. As indicated, the plate current reversed direction. A single primary electron striking a metal surface may

free up to 8 or 10 secondary electrons for good surfaces. The plate in modern tetrodes is treated to reduce secondary emission and the effects shown in Fig. 104 are not so pronounced. Because of the erratic performance in the region of points 1 to 4, tetrodes ordinarily are not operated in this region, but are operated over the straight portion of the curve from point 4 to point 5 and beyond. The tetrodes that have been described in this section often are called **screen-grid tubes.** These tubes once were used extensively in radio-receiving sets, but have been replaced to a considerable extent by the tubes to be described in the next section.

The Pentode.—As mentioned in the preceding section, tetrodes usually are operated over the straight portion of the characteristic curve from point 4 to point 5 and beyond. This ensures that the voltage from plate to cathode never falls below the voltage from screen grid to cathode, that the secondary electrons are drawn back to the plate, and that the effects of secondary emission do not influence the characteristic curves. Because of this, the plate voltage changes, when the tube is amplifying a signal, must never cause the plate voltage to fall below point 4.

By adding a third grid, called a **suppressor grid,** between the screen grid and plate, the *effects* of secondary emission are eliminated. Such a tube is a **pentode,** or **five-electrode tube.** This third grid also further shields the control grid from the plate and thus further reduces the tendency to feed back and cause undesired oscillations when the tube is used as an amplifier. The presence of the third grid also increases the amplification factor and plate resistance.

Note the wording in the preceding paragraph: *The effects of secondary emission are eliminated.* In the pentode, high-speed electrons strike the positive plate, and thus secondary emission still occurs. The suppressor grid keeps the secondary electrons from flowing to the screen grid and thus the characteristic curves are as shown in Fig. 105.

As previously stated, the suppressor grid is between the plate and screen grid. The suppressor grid is connected *directly* to the cathode and is at zero (cathode) potential. For this reason, when the secondary electrons are knocked from the plate, they find themselves influenced by one positive electrode only—the plate—and, hence, they all return to the plate. Because of this action, region 1 to 4 of Fig. 104 is eliminated, and the smooth curves of

Fig. 105 result. In pentodes, therefore, the plate potential need not always be higher than the screen-grid potential.

An examination of Fig. 105 will show that the grid has about the same effectiveness as in the tetrode. Because of the addition of the suppressor grid, the plate is further shielded from the electrons in the space-charge region, and is less effective than before, as indicated by the flatness of the plate voltage-plate current curve. For the plate to influence the electrons it collects, the effect of the positive plate would have to "reach through" the suppressor grid, the screen grid, and the control grid; to do so would be almost im-

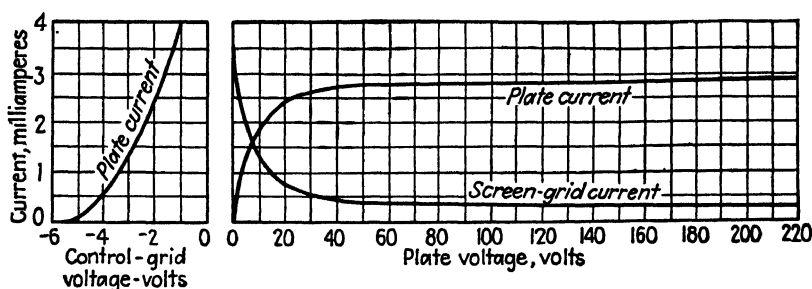



FIG. 105.—Curves for a small suppressor-grid pentode. The suppressor grid eliminates the undesired *effect* of secondary emission, giving the smooth curves shown as compared with the curves of the tetrode of Fig. 104.

possible. The plate of the pentode receives those electrons which the positive screen grid pulls through the negative control grid, and which have velocities sufficient to carry them through the suppressor grid so that the plate can collect them. As a result of the inability of the plate to affect the electrons, the amplification factor is very high, of the order of 1,000 or more; and the plate resistance is very high, of the order of 1,000,000 ohms or more. The mutual conductance is about the same as for the other tubes considered, about 1000 micromhos. The static characteristics of a pentode are determined in a circuit such as Fig. 103. The construction of a typical pentode is shown in Fig. 106.

Static and Dynamic Characteristics.—The curves of Figs. 101, 104, and 105 are known as **static curves,** because they are taken with direct potentials on the electrodes and no alternating signal voltages or currents are being passed by the tube. The static curves are very useful for explaining the electronic nature of tube performance. These curves are of value in determining tube coefficients as explained under Fig. 101.

The data in this figure are for a triode, and the coefficients are easily determined. But for tetrodes and pentodes, the plate voltage-plate current curves (Figs. 104 and 105) are so flat that the graphic methods of Fig. 101 are only approximations.

In practice dynamic methods are used to determine tube coefficients and to predict tube operation. For such tests, alternating

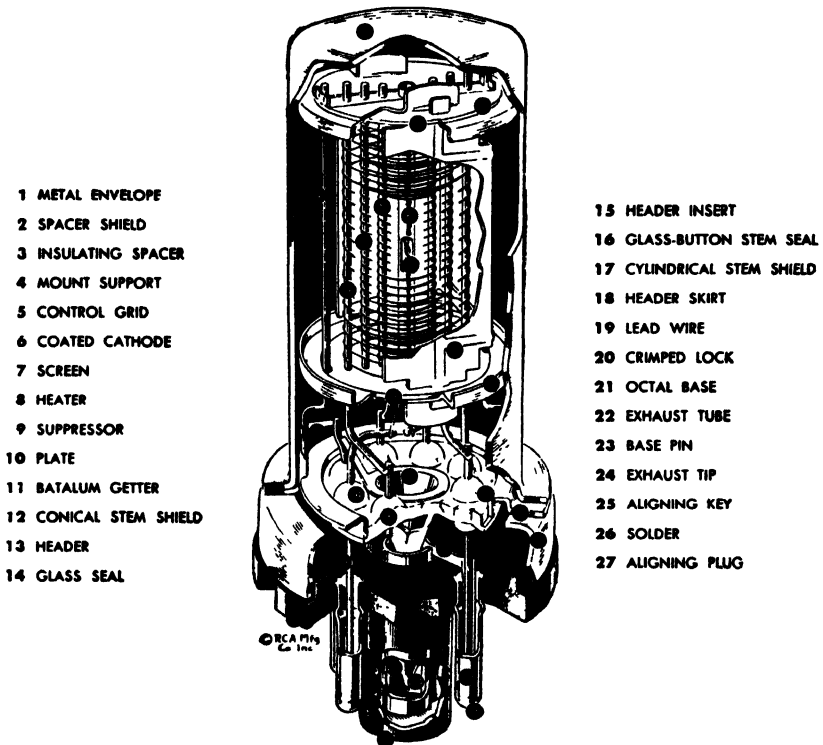



FIG. 106.—Details of construction of a small pentode. (*Radio Corporation of America.*)

signal voltages are impressed on the control grid. These test methods have been devised with much care and are explained in Standards of Electronics—Methods of Testing Vacuum Tubes, a pamphlet issued by The Institute of Radio Engineers. These standards fully explain the laboratory methods to be followed in tube tests.

Remote Cutoff Tubes.—An examination of Figs. 101, 104, and 105 will show that the grid voltage-plate current curves approximate straight lines until the plate currents reach low values, when

the tubes abruptly cut off and the plate current falls to zero. If it is assumed that these curves are straight lines, then it follows that the amplification factor is the same, irrespective of the grid-bias voltage and the plate current (Figs. 101 and 102). Thus when a tube of this type is used as an amplifier (Fig. 100), there can be only one value of amplification factor.

In the design of radio receivers a need exists for a tube that has many possible values of amplification, from low values to high values. This is because of **fading** (page 521) of the signal from distant radio transmitting stations. If such a tube were available, then a circuit could be arranged so that when the received radio signals were weak, the amplification of the tube could be large, and when the signals received were strong, the amplification of the tube could be small. In this way **automatic-volume control** could be accomplished, and the output of the radio-receiving set could be made independent of fading and would remain constant, at least over a wide range of received signal strength.

Strictly speaking, in the vicinity of cutoff the amplification factor of a tube varies over wide limits. But, the curvature is such that to prevent excessive distortion, infinitesimal voltages only could be impressed. (Study Fig. 100, and ascertain the shape of the alternating component of the plate current if the grid voltage-plate current characteristic were curved as in the vicinity of cutoff of Figs. 101, 104, and 105.)

To provide the variable amplification factor, and to make possible the use of "large" applied signal voltages (and not infinitesimal values), the **remote-cutoff tube** was developed. A comparison of the characteristics of this tube with those of a conventional screen-grid tetrode are shown in Fig. 107. There are many points at which the remote-cutoff tube may be operated. As indicated in Fig. 108, different values of amplification factor result. The use of this principle will be explained further on page 432.

This remote cutoff may be achieved in many ways. A common method is to omit one or more turns at the center of the spiral of wire constituting the control grid. With this type of grid, cutoff is gradual. As the control grid is made progressively more negative, cutoff occurs first at portions of the tube structure where control-grid wires exist. This reduces the current to some extent. As the grid is made more and more negative, a voltage bias value is finally reached, at about -50 volts for a typical tube, where the

current to the plate is forced to zero. At this voltage, the gap in the control-grid wires has been effectively closed by the strong electric field existing at −50 volts. The terms **variable-mu tube** and **supercontrol tube** are often applied to this tube.

Multigrid Tubes.—To the inexperienced, the vast array of tubes on the market is bewildering. Even the experienced person often has difficulty in classifying a given tube. To assist in clarifying the situation, the following paragraphs are included.

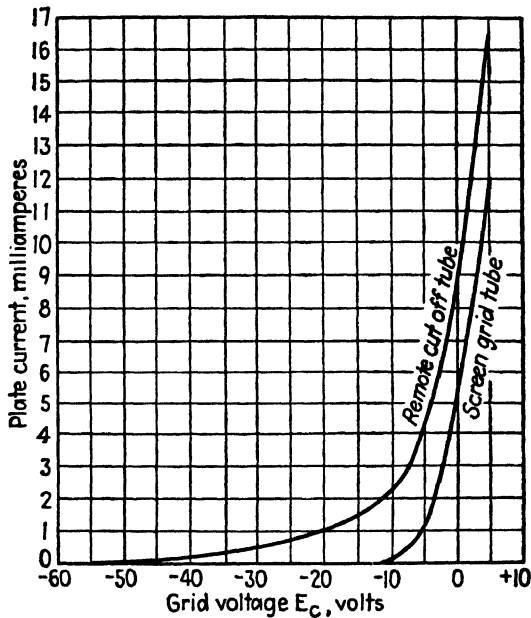



FIG. 107.—A comparison of the characteristics of a remote-cutoff or supercontrol tube with those of an ordinary tetrode.

Multigrid, or multielectrode, tubes usually are classified as tubes containing more electrodes than a triode. There are two classes of multigrid or multielectrode tubes: (*a*) those tubes designed to perform some function that cannot be performed by a triode and (*b*) those tubes having additional electrodes to perform simultaneously more than one function.

The tetrode and pentode are examples of the first class. They will perform excellently as radio-frequency amplifiers, but the triode will not. The "duplex-diode high-mu triode" is an example of the second class. This composite tube performs simultaneously

more than one function. It consists essentially of *two* tubes in the same envelope. One part (duplex diode) can be used to demodulate (or detect) the radio-frequency signal, producing an audio-frequency component, and the other part (the high-mu triode) can be used to amplify this audio component.

In studying vacuum tubes, it is well to remember that there are four basic types of tubes, diodes, triodes, tetrodes, and pentodes. In general, all vacuum tubes (such as used in radio receivers) are of these types and perform like these types, but sometimes two units are combined in one envelope.

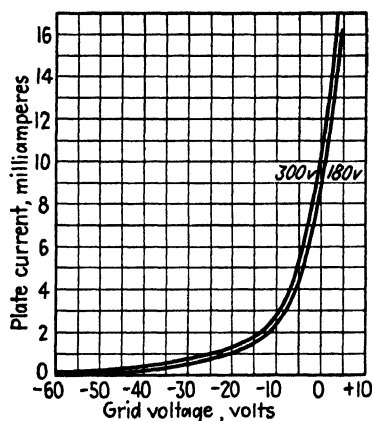



FIG. 108.—Curves for the remote-cutoff tube of Fig. 107 at two plate voltages. Using the method of Fig. 101, at 8 milliamperes the amplification factor is approximately (300 − 180)/(1.25 − 0) = 96. At a plate current of 0.5 milliampere the amplification factor is approximately (300 − 180)/(35 − 29) = 20. This shows that the amplification factor varies with the magnitude of the grid-bias voltage.

Voltage-amplifying and Power-output Tubes.—It becomes very important in a study of radio to have clearly in mind the fundamental purpose for which vacuum tubes, such as triodes, tetrodes, and pentodes, are used. In general, there are two separate and distinct purposes for which tubes are used, although sometimes the distinctions are not obvious. These two basic uses of tubes can be illustrated by the equipment used in radio transmission and reception.

In a radio-transmitting set, the feeble signal voltage from a microphone is amplified in **vacuum-tube voltage amplifiers,** which contain **voltage-amplifying tubes,** until the voltage is sufficient to drive a **vacuum-tube power amplifier** which contains **power-output tubes,** or **power tubes.** When the signal power is of sufficient strength (the amount depending on the type of transmitter), the signal is used to modulate the carrier, and the output is again (in some transmitters) amplified in power amplifiers containing power-output tubes that furnish power to the transmitting antenna.

In an elementary radio-receiving set the feeble signal voltage received by the antenna is amplified by voltage amplifiers and the signal is then demodulated or detected. The resulting audio-

frequency component is then amplified in voltage amplifiers, using voltage-amplifying tubes, and this strengthened signal is used to drive a power amplifier which contains power-output tubes which drive the loudspeaker.

From the standpoint of the electronic operation and the basic characteristics, voltage-amplifying tubes and power-output tubes may be said to be similar, or in fact, the same. The essential difference is this: Tubes designed primarily for voltage amplification are small physically and have internal construction (such as size and spacing of the grid wires) that make them essentially low-current devices having high amplification factors and plate resistances.

By contrast, power tubes are larger physically (usually) than voltage-amplifying tubes, and have large electrode structures with rugged grids and large spacings between the grid wires. These tubes readily pass comparatively large currents and will furnish large amounts of power to loudspeakers, transmitting antennas, etc. Power tubes have low amplification factors and low plate resistances compared with voltage-amplifying tubes.

Voltage-amplifying tubes include triodes, tetrodes, and pentodes. Power-output tubes include triodes and pentodes. Power-output tetrodes have been developed and are used to some extent, but never have proved too popular. The tubes that have been designated as power-output, or power, tubes sometimes are called **power-amplifier tubes.** Although this term is used, it is pointed out that in many applications the grid-signal power input essentially is zero, and, hence, the tubes are not power *amplifiers* in the exact sense.

Beam-power Tube.—Certain special tubes have been devised and are in use that are not exactly like the tubes described thus far in this chapter; yet these special tubes perform like one of the basic types of tubes. The beam-power tube is an example of a special tube which has achieved wide use, and which continues to grow in popularity.

The beam-power tube is a power-output tube that has four electrodes but functions like a pentode. This is possible because the suppressor grid is omitted, yet the undesired effects of secondary emission are eliminated by other means. To understand the principle of the beam-power tube, it is well to review briefly the theory of the pentode.

In the pentode the secondary electrons are kept from flowing to the screen grid by interposing a region of zero potential between the plate and screen grid. Then the secondary electrons are attracted only by the positive plate and, hence, return to it. This region of zero potential is established by placing the suppressor grid between the plate and screen grid and connecting the suppressor grid directly to the cathode, which is at zero potential. Now the suppressor grid is placed there only to establish a region of

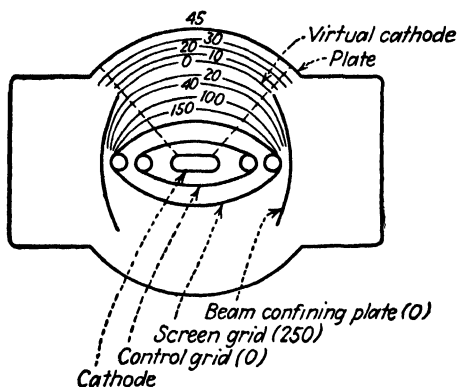



FIG. 109.—Top view of electrodes and the voltage distribution when the plate is at low potential. The zero-potential region forms a virtual cathode, and the electrons reaching the plate appear to come from this virtual cathode. The zero-potential region keeps the screen grid from attracting the secondary electrons emitted by the plate. In this way the tube behaves like a pentode but does not equire a suppressor grid. (*Radio Corporation of America.*)

zero potential, and if such a region could be established *by other means*, then secondary emission effects would be controlled as in a pentode.

This is accomplished in a beam-power tube as shown in Fig. 109. The electrodes and the "beam-confining plate" are so shaped that the electric field and voltage distribution within the tube essentially are as if the beam-confining plate were continuous instead of having a gap in it. Because of these shapes, secondary electrons from the plate find a region of zero potential (which offers no attraction) between them and the screen grid, and these secondary electrons return to the positive plate. Thus the tube performs essentially as a pentode; yet it has only two grids.

The voltages indicated in Fig. 109 represent conditions when the tube is in operation and are for the part of the signal cycle when the plate voltage has fallen to the minimum value (page 274). As in-

dicated, the plate is at 205 volts less potential than the screen grid, and yet the region of zero voltage called a "virtual cathode" exists. This region is called a "virtual cathode" because, as far as the plate can tell, all electrons appear to come to it from a cathode located along the zero line.

Various refinements have been incorporated in the beam-power tube. For instance, the control-grid and screen-grid wires are

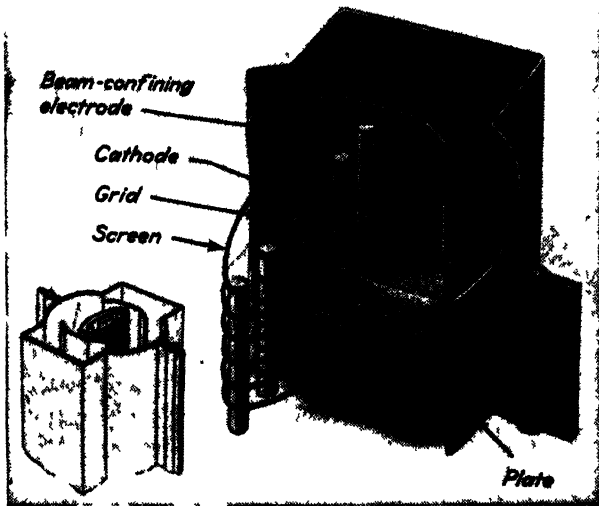



Fig 110.—Electrode structure in a small beam-power tube. Note how the negative control grid forms the electron path so that relatively few electrons strike the screen-grid wires. (*Radio Corporation of America.*)

aligned so that the positive screen-grid wires are directly behind the negative control-grid wires. In this way the negative control grid in repelling the electrons also assists in keeping these electrons from striking the positive screen-grid wires immediately behind as shown in Fig. 110. The screen-grid current is low because of this action, and, hence, the over-all efficiency is increased over that of a comparable pentode. The characteristics of a beam-power tube are similar to those of a power-output pentode.

The Gas Diode.—The tubes that have been considered in the preceding pages of this chapter were of the high-vacuum type, and were evacuated to such a degree that their operation essentially was unaffected by the presence of the remaining residual gas. The so-called "gas tubes," to be considered in the following sections, are first evacuated to a high degree to remove undesired gases, and

then a small amount of a suitable gas is admitted and the tube is sealed. The gases used are the very stable chemically inert gases, such as helium, neon, and argon. In certain gas tubes, known as **mercury-vapor tubes,** a small amount of liquid mercury is placed in the evacuated tube. At the low pressure some of the mercury vaporizes, filling the tube with mercury vapor which has the properties of a stable gas.

A gas diode consists of a heated cathode, an anode, and the gas, or mercury vapor, in a glass or metal container. The cathode may be of either the filament type or the indirectly heated type. The gas (or vapor) is in the tube to reduce the effect of the negative space charge and thus decrease the voltage drop across the tube. Such tubes are used extensively in rectifiers (Chap. VII) to change from alternating-current power to direct-current power. Since these may be rectifying large amounts of power, high efficiency often is of much importance. Other factors being comparable, the tube with the least voltage drop between cathode and anode will be the most efficient rectifier.

The gas (or vapor) in a tube reduces the voltage drop across the tube by neutralizing the negative space charge that exists (largely) as a sheath about the cathode. As explained on page 177, this sheath is composed of negative electrons thrown off by the cathode. The current, consisting of negative electrons, that flows from the cathode to the anode must pass through the negative space charge and is opposed by it. The positive plate must, in a sense, pull the negative electrons through the negative space charge, and this requires a voltage of, perhaps, +50 volts, depending on the tube and current magnitude. If some way could be devised to *neutralize* the opposing *effect* of the space charge (without destroying it because it acts as a reservoir of electrons), then the plate voltage, which is the voltage drop across the tube, would be much less, and the tube would be more efficient as a rectifier.

The way in which a gas neutralizes the space charge can be explained by Fig. 111. Each of the figures contains a cathode and an anode. The heavy black dot, when added to the graphical symbol of a tube, indicates that the tube contains gas or mercury vapor. The first figure shows the condition in the tube when the cathode is hot and a space charge is around it, but with *no positive potential applied to the plate.* Under these conditions the gas atoms are

essentially unaffected and they are electrically neutral as indicated by the absence of − or + signs.

The second drawing of Fig. 111 indicates conditions within the gas tube after the anode is made positive with respect to the cathode. At the instant that the positive plate voltage was applied, negative electrons were drawn from the negative space-charge region over toward the anode, or plate, and many of these electrons achieve high velocities. On their way to the positive plate, some

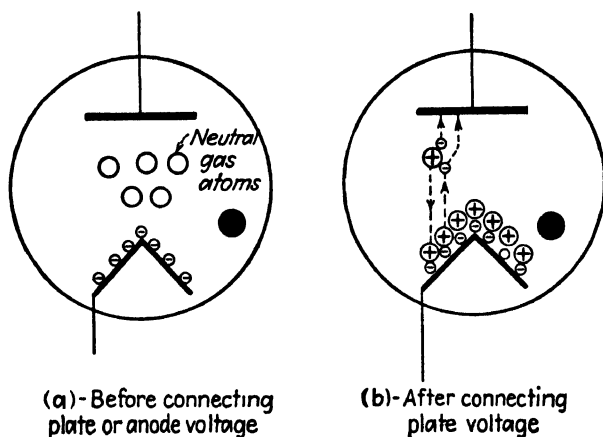



Fig 111.—In a gas tube the positive ions formed by collision neutralize the negative space charge The large black dot in the graphical symbol for a tube indicates that it is a gas tube

of these high-speed electrons strike neutral gas atoms, and in the gas tubes under consideration these collisions result in **ionization.**

An atom of gas is conveniently thought of as being composed of a positive heavy central nucleus about which negative electrons are revolving. There are enough revolving negative electrons to equal the positive charge on the nucleus, and the atom as a whole is neutral. When **ionization by collision** occurs, the high-speed electrons tear away one or more of the revolving electrons and the atom no longer is neutral. It has been broken into two charged particles called **ions.** One ion will be the negative electron which was torn away and the other ion will be a **positive ion** which is the remainder of the atom, that is, the atom minus one or more electrons. The **negative ion** (electron) has little mass and will dart quickly to the positive anode. The positive ion will have a mass

essentially that of the gas atom, and will move slowly toward the negative cathode. Because of their slow motion, the massive positive ions accumulate within the tube and neutralize the negative space charge.

The process is illustrated in Fig. 111*b*. A small negative electron on its way to the positive anode has collided with what was originally a neutral gas atom and has ionized it by knocking out a negative electron. As indicated, this electron darts to the positive

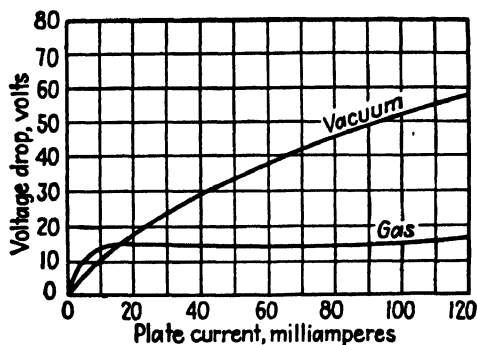



FIG. 112.—In the gas diode the internal voltage drop between cathode and plate is much less than for a comparable high-vacuum diode. Also, the voltage drop in the gas diode is largely independent of the current flow over the normal operating range. This is because the ionization by collision and the resulting positive ions largely neutralize the negative space charge.

anode, leaving the massive position ion (atom minus an electron), which travels slowly toward the negative cathode. The accumulation of slow-moving positive ions quite effectively neutralizes the negative space charge. As a result of this, the voltage drop across the tube (that is, the voltage required between the cathode and anode to cause plate-current flow) is much less than if the gas and the positive ions were not present. In fact, the voltage need be but little larger than that required to give the electrons energy sufficient for ionization. For typical tubes this voltage drop is about 15 volts. Furthermore, it is largely independent of the magnitude of the current over the operating range, as Fig. 112 indicates. The greater the current, the more the ionization, and the better the space charge is neutralized.

In the operation of gas tubes, the cathode should always be at correct operating temperature before the anode (plate) voltage is applied. This is to ensure that the negative space charge has

built up and that there is a sheath of negative electrons about the cathode. This sheath acts somewhat as a "buffer" for the positive ions moving toward the cathode. If the negative space charge is not present, or if the voltage between anode and cathode is too great, then the massive positive ions may have sufficient energy to damage or to ruin the cathode completely.

As was previously mentioned, both filament and indirectly heated cathodes are used in gas tubes. These are oxide coated to supply the large number of electrons that are required, gas tubes being used to a great extent for rectification or other purposes where large currents flow. If the voltage drop across a gas tube with an oxide-coated cathode exceeds about 22 volts, then the positive ions acquire sufficient energy to damage the cathode seriously. This voltage may be reached and exceeded in rectifier circuits if (*a*) the space charge is too low because of insufficient cathode heating current or insufficient heating time, and (*b*) if the cathode-to-anode electron current is too great. For the tube of Fig. 112, it is noted that at high values of current the voltage drop is increasing. If the current through the tube is made progressively greater, this drop will increase until the cathode is ruined and perhaps the tube arcs across. Time-delay relays often are used in equipment employing gas tubes, and these relays apply the anode voltage *after* the cathode has had sufficient heating time. Although such precautions may be overlooked with small tubes having filament-type cathodes, which heat rapidly, they must not be disregarded with large gas tubes.

Somewhat complicated cathode structures are used in large *gas* tubes, such as the ones employed for rectifying the large amounts of power for the power tubes of radio transmitters. Of course these cathodes consist of heaters and oxide-coated surfaces, but the structures are complicated with heat shields that reduce energy loss through infrared light radiation and otherwise increase the thermal efficiency of the cathode. These cathodes have high thermal storage capacity and their temperature increases slowly. For instance, one widely used mercury-vapor rectifier tube requires 42 amperes at 5 volts and a heating time of 5 minutes. The drop across the tube is approximately 15 volts, and it will pass a direct current of 5 amperes, and can be used in rectifiers with outputs of about 10,000 volts.

The Gas Triode.—If a grid is placed in a gas tube, it might be

expected that it could control the plate current much as in a high-vacuum triode, but this is not the case. The grid in the gas triode can control the start of current flow through the tube, but once the current starts, *the grid loses control,* and cannot again control the current until the tube is deenergized. After this has been done, the grid can once again control the start of current flow, but again loses control once current flows through the tube. The reason for this peculiar action will now be explained, using Fig. 113.

The **gas triode** (often called by the trade name **Thyratron**) consists of an oxide-coated cathode, a plate, or anode, and a grid in an evacuated glass, or metal, envelope that contains a small amount of gas, or mercury vapor. From Fig. 113 it would be inferred that the grid is a simple structure between the cathode and anode. In some gas triodes this is true, but in many the grid is a very complicated structure, perhaps surrounding the cathode and anode. The graphic symbol for a gas triode is shown in Fig. 113. The oxide-coated cathode may be a filament, it may be an indirectly heated oxide-coated cylinder, or it may be one of the complicated heat-shielded structures discussed in the preceding section.

FIG 113—Graphic symbol and circuit for testing the grid-controlled gas-filled tube or Thyratron The resistors R_a in the anode or plate circuit and R_g in the grid circuit are protective resistors. Switch S can be opened to stop anode current flow and permit deionization.

The circuit of Fig. 113 can be used to study the gas triode. Suppose that a gas triode of the type used in cathode-ray tube sweep circuits (page 576) is under test. If the voltage dividers are adjusted so that the grid is, say, 15 volts *negative* and the anode is 100 volts *positive*, the tube will not conduct, because -15 volts on the grid is sufficient to overcome the effect of $+100$ volts on the anode. The grid voltage of -15 volts repels the electrons and keeps current from flowing to the positive anode. Thus far, the gas triode has functioned like a high-vacuum triode.

Now suppose that with the grid maintained at -15 volts the anode is made more and more positive. At an anode voltage of about $+140$ volts the grid no longer can hold back the electrons,

electron current flows to the anode, ionization by collision occurs, and the grid loses control. The grid loses control for this reason: When ionization by collision occurs, the negative grid attracts the positive ions and repels the negative ions (electrons). The electrons quickly dart out of the space near the grid, leaving the massive positive ions. Thus a positive space-charge sheath builds up around the grid, this positive sheath neutralizes the effect of the negative grid, and the grid loses all control. The tube then operates essentially as a gas diode, and current flows through the tube. Thus the grid in a gas triode can control the start of current flow, but once current flows it loses all control. It cannot regain control until the anode voltage falls to zero, current flow stops, and **deionization** occurs. By deionization is meant the neutralization and **recombination** of the positive and negative ions and the return of the ions to normal uncharged gas atoms. A certain time interval is required for deionization to occur.

FIG 114 —Characteristics of a small grid-controlled gas-filled triode or Thyratron. This tube contained argon gas

The characteristics of a small gas triode are shown in Fig. 114. This is a plot of the critical grid voltages and corresponding anode voltages at which the grid loses control and current flows through the tube. The various points are obtained as explained in the preceding paragraph. It is of particular importance to note that the *magnitude* of the current that flows through the gas triode must be regulated by the resistor R_a in the plate, or anode circuit. Otherwise the tube will arc across. A current-limiting resistor R_g also must be placed in the grid circuit to keep the grid current (it attracts positive ions) within operating limits.

The Gas Tetrode.—There is only one operating curve for a gas triode, Fig. 114 being an example. The operating characteristics of a given gas triode cannot be changed at will to fit given requirements. In the gas tetrode, however, the characteristics can be changed, and the tube is more flexible in its use.

If a second grid is added to the gas triode, making the tube a

gas tetrode, then the potential of the second grid can be made to change the breakdown points. This second grid usually is called a **shield grid,** and sometimes a screen grid, but it does not function as in the high-vacuum tetrode (or screen-grid tube).

The second grid in a gas tetrode has the ability to shift the breakdown characteristic curve of Fig. 114 to the left or to the right, thus providing a variety of characteristic curves depending on the potential of this second grid. If the second grid is made a few volts *positive* with respect to the cathode, then the curve of Fig. 114 is shifted to the left because the first (control) grid must be maintained more negative to control the initial operation (up to the point of breakdown). If, on the other hand, the second grid is made slightly negative, then it assists the control grid, and less control-grid voltage is required to determine the point at which current flows. In fact if the second grid is made sufficiently negative, it becomes necessary to impress a positive potential on the control grid. If this is not done, then the tube cannot be made to conduct over the usual range of positive anode values.

FIG. 115.—Circuit for determining the characteristics of a cold-cathode two-electrode tube. The black dot indicates that the tube contains gas.

Cold-cathode Two-electrode Gas Tubes.—These usually consist of a glass tube or bulb filled with an inert gas, such as neon or argon, at reduced pressure. There is no hot cathode. The two electrodes are of various shapes, depending on the purposes for which the tubes are to be used. Often the electrodes are identical and either may serve as a cathode. The so-called "neon" tubes used in advertising are cold-cathode tubes; these will not be discussed. Only the type used in communication will be considered.

A circuit for testing a cold-cathode two-electrode gas tube is shown in Fig. 115. The circular electrode is the cold cathode, and the black dot indicates that the device contains gas. The plate or anode is represented by the usual symbol. To obtain the characteristics, voltage divider R_1 is set to give zero volts, and rheostat R_2 is adjusted for maximum resistance. The voltage divider then is varied and the voltage increased until the tube breaks down and conducts as at point 1 of Fig. 116. As soon as the tube conducts, the voltage drop across it falls to point 2. The curve from point

2 to point 3 is obtained by varying R_3 and reading the voltage drop at given current values.

The initial breakdown at point 1 is explained as follows: In any body of gas the *majority* of the gas atoms contain equal positive and negative charges and, hence, are electrically neutral. There are, however, always a few charged ions that are produced by natural causes, perhaps by "stray" radiations. When the voltage is applied between the cathode and anode, these charged ions are

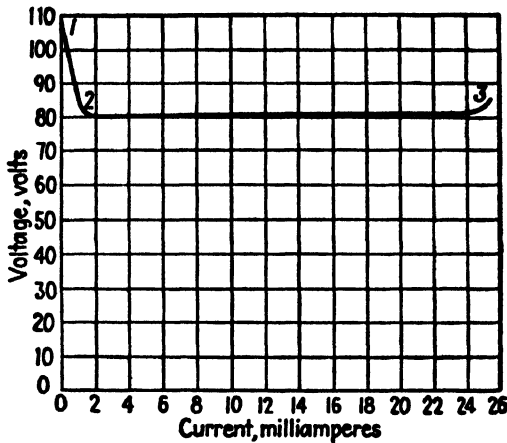



FIG. 116.—For a cold-cathode two-electrode gas-filled tube the voltage across the tube remains substantially constant for a wide range of current values. It is this fact that makes the tube useful as a voltage regulator. The portion of the curve between points 1 and 2 is approximate.

attracted to the electrodes. As these ions progress toward their respective electrodes, they gather momentum and produce ionization by collision (page 201). This produces more ions, which in turn causes more ionization, and so on, until the gas becomes a good conductor. The current is then limited by rheostat R_3, which keeps the tube from arcing across.

Electrons from the ionized gas are attracted by the positive plate, flow on around the external circuit, and then to the cathode. They must leave the cold cathode and return to the gas. Just how can they do this? The answer is not so simple as for a thermionic cathode, where it is assumed that the heat energy causes their liberation. In a cold-cathode tube it is thought that the positive ions cause the electrons to escape from the cold metal cathode. It is probable that the slow-moving positive ions ac-

cumulate very near the cathode surface, producing a sheath of positive electricity, and that these positive ions produce a strong electric field at the cathode surface. This assists the electrons in escaping from the cold metal (page 173). Also, positive ions falling into the cathode surface probably assist in freeing electrons. Of course the temperature of the cathode rises above room temperature, but not to a point where the emission is thermionic in nature as in the case of hot cathodes.

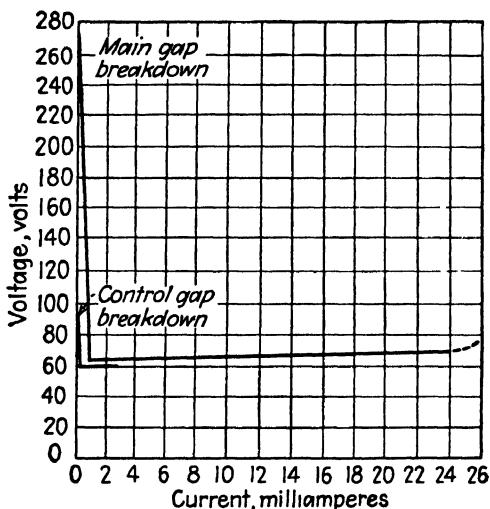



FIG. 117.—Characteristics of the control gap and the main gap for a cold-cathode three-electrode gas-filled tube.

If the two electrodes are small parallel disks, or similar electrodes such that the phenomena can be observed, a small area of the cathode will be seen to glow as soon as current flows through the tube (point 2, Fig. 116). If rheostat R_2 is varied so that more and more current flows, the glow will become larger and larger until point 3 is reached and the cathode is all covered with the glow. Over the region 1–2, the voltage drop remains essentially constant as the current is increased, which means that the internal resistance of the cold-cathode tube must decrease. These interesting properties are very useful in radio apparatus, and these tubes are often called **voltage-regulator tubes** (page 250).

Cold-cathode Three-electrode Gas Tubes.—In the two-electrode tube just considered, there was *one* gap across which the discharge

occurred. In the three-electrode cold-cathode tube the arrangement is such that there are *two* gaps across which discharges may normally occur. A variety of electrode shapes are used, but in any tube there is a **control gap** between the **starter,** or **control anode,** and cathode, and a **main gap** between the **main anode** and cathode. The characteristics of a typical cold-cathode three-electrode gas tube are shown in Fig. 117. These characteristics are obtained by using the circuit of Fig. 115 and testing each gap separately. The electrode not in use should be connected as specified by the manufacturer.

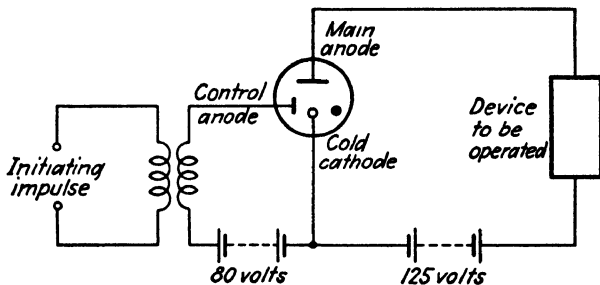



FIG. 118.—Showing how a cold-cathode three-electrode gas tube can be used as a relay. If the tube has the characteristics shown in Fig. 117, it will not conduct when the potentials are as shown in the diagram. An initiating impulse of a few volts will, however, cause the control-anode gap to break down, and then the main gap will break down, thus operating the device at the right.

As Fig. 117 indicates, the control gap breaks down at a voltage much lower than that required by the main gap. After breakdown, the gaps perform as explained in the preceding section. This tube performs as an electrical relay, as shown in Fig. 118. For the tube whose characteristics are shown in Fig. 117, the voltages between the anodes and cathode should be somewhat as indicated. Neither gap will break down, and essentially no current will be drawn from the sources. If now an initiating impulse of a few volts positive is impressed as indicated, the starting or control anode will be driven sufficiently positive, and breakdown of the starting gap will occur. This will fill the tube with ions, and the main gap will conduct, causing operation of the device in the main anode circuit. Thus this tube can operate as a relay because a small impulse can control a much larger output. It is used also as a voltage regulator.

Phototubes.—As mentioned on page 174, the energy of a beam of

light can release electrons from metals. This principle is used in phototubes. These consist (Fig. 119) of a light-sensitive cylindrical cathode and a centrally located straight wire serving as an anode. The glass bulb may be evacuated, or may contain inert gas. The phototubes in common use with incandescent electric-light sources have cathodes covered with a cesium-oxygen-silver layer that is an excellent emitter for such sources. The number of electrons emitted is directly proportional to the intensity of the light striking the phototube. A circuit for testing a phototube is shown in Fig. 120.

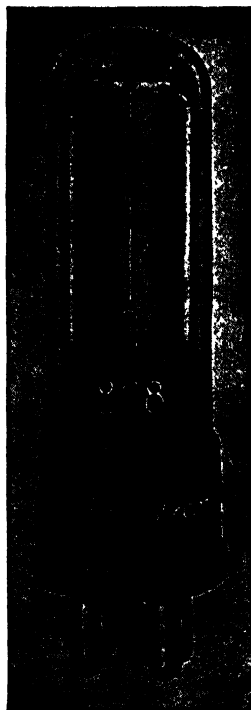



FIG. 119.—A phototube. (*Radio Corporation of America.*)

The High-vacuum Phototube.—If the illumination at the phototube cathode is held constant at 0.1 lumen and the voltage is varied, the curve of Fig. 121*a* results. The tube "saturates" and the current becomes constant because the anode is taking the electrons as rapidly as they are released by the light striking the cathode. If the illumination is increased to 0.5 lumen, a larger electron current flows to the positive anode, because the increased illumination releases more electrons. If the data are plotted as in Fig. 121*b*, a straight line results, indicating that the electrons emitted and the current output are directly proportional to the intensity of illumination. The number of lumens can be determined from the relation

$$F = \frac{AC}{d^2}, \tag{74}$$

where F is in lumens, when A is the projected area of the cathode (not the total surface area, but the product of diameter and length), C is the candle power of the light source, and d is the distance from the cathode to the light source, measured in the same units as A. For accuracy, tests should be made in accordance with the Standards of Electronics (page 193). Approximate tests can be made using ordinary light meters to measure the illumination. The

high-vacuum phototube is a very stable device, but has a low output as Fig. 121 indicates.

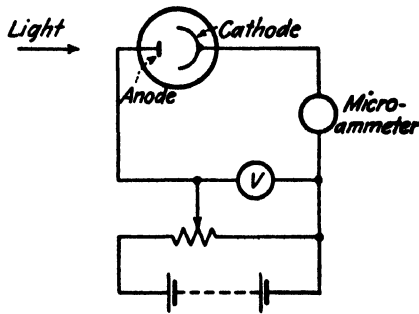



FIG. 120.—Circuit for testing a phototube. When a gas tube is tested, a current-limiting resistor should be placed in the anode or plate circuit to protect the tube.

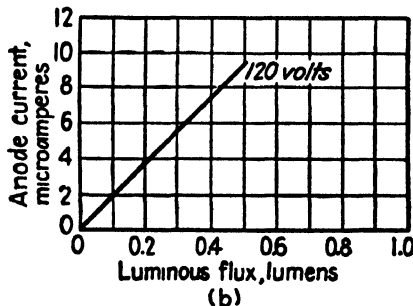
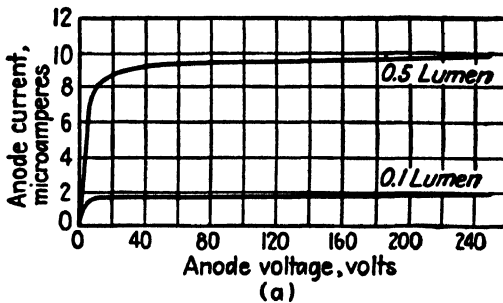



FIG. 121.—Characteristic curves of a high-vacuum phototube. The curve of figure (*b*) proves that the output current of a phototube varies directly with the light striking the cathode. The intensity of illumination is measured in lumens and is found by the equation $F = AC/d^2$, where A is the effective area of the cathode (that is, the projected area), C is the light intensity of the source in candle power, and d is the distance between light source and cathode.

The Gas Phototube.—If a small amount of an inert gas, such as argon, is placed in a phototube, the output sensitivity is increased

greatly. This is because the emitted electrons as they move toward the positive anode cause ionization by collision, and this increases the current flow. Because the gas phototube is more sensitive than the high-vacuum type, the gas phototube is the more widely used, although it is less stable and in certain applications causes some distortion.

The characteristics of a typical gas phototube are shown in Fig. 122. The curves differ from those of Fig. 121 because of the presence of the gas. The two tubes otherwise are identical. In the

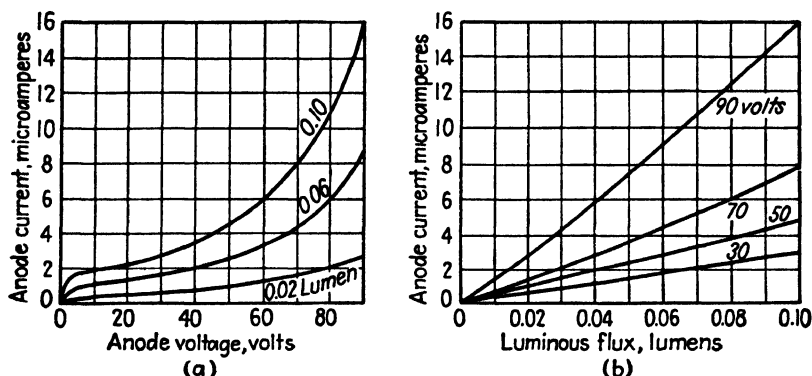



FIG. 122.—The characteristics of a gas phototube. These curves should be compared with Fig. 121.

gas phototube the greater the voltage, the greater the ionization—and hence the greater the current through the tube. As indicated, the current increases rapidly at the higher voltages. Care must be exercised to hold the current within prescribed limits. A series resistance of perhaps 1 megohm is often placed in series with the anode to limit the current flow and to prevent the tube from arcing across. That the phototube current essentially is directly proportional to the illumination is indicated by Fig. 122*b*.

Multiplier Phototubes.—The output of a phototube is very small, and vacuum tubes have been used extensively to increase the output so that a phototube could be used directly to operate relays, counters, etc. A phototube with a secondary-emission multiplier in the same bulb now is available (Fig. 123). The various electrodes are interleaved and geometrically arranged so that the electron path is from one electrode to the other. Each electrode is at a higher positive voltage than the preceding electrode. Then,

the electrons emitted from the photosensitive electrode by the light striking it are attracted to the next electrode, where secondary emission occurs, and the number of electrons is increased by some factor less than 10. These then go to the next electrode, where secondary emission again occurs, and the number of electrons again is increased. By this action, the final electron current is increased many times over the original photoelectric current.

The Photovoltaic Cell.—A review of the preceding section will show that phototubes act merely as variable high resistances or as

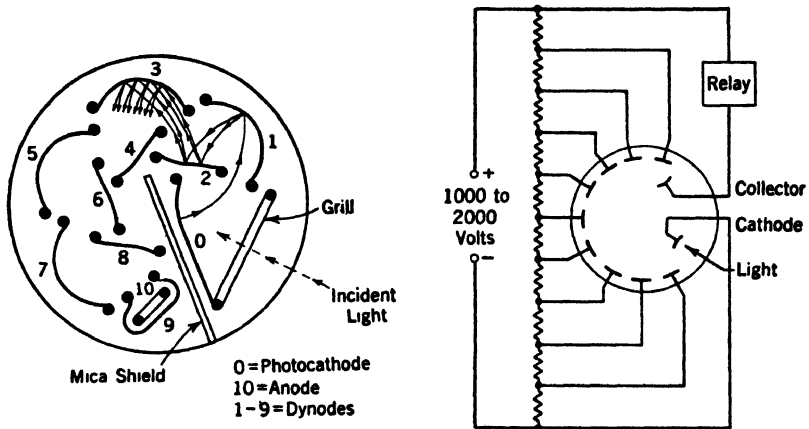



FIG. 123.—Arrangement of electrodes and typical circuit for a multiplier-type phototube. The relay indicated is the device to be operated. (*Radio Corporation of America.*)

light-controlled electric valves. A voltage must be impressed between the cathode and anode, and the tube allows this voltage to force more or less current through the circuit, depending on the intensity of illumination.

There is a photoelectric device that does not need an external applied voltage. This device is called a **photocell**, or perhaps more correctly a **photovoltaic cell**, and it generates its own voltage and will itself force current through an external circuit. The radiant energy of the illumination is converted into electric energy, and this causes the current flow. Since no external source (such as a battery) need be used, the photocell is well adapted to operation at remote points, one use being to turn radio-antenna warning lights on and off during the 24-hour period.

The photocell consists essentially of a metal disk on which a layer

of light-sensitive material has been formed or deposited. In the cell most commonly used, the disk is of iron and the layer is a selenium compound. A *very* thin layer of conducting material is deposited on top of the light-sensitive layer. The conducting layer must be so thin that light can readily pass through and enter the light-sensitive layer beneath. The thin conducting layer forms one electrode and spreads the current out over the light-sensitive material. The iron disk is the other electrode.

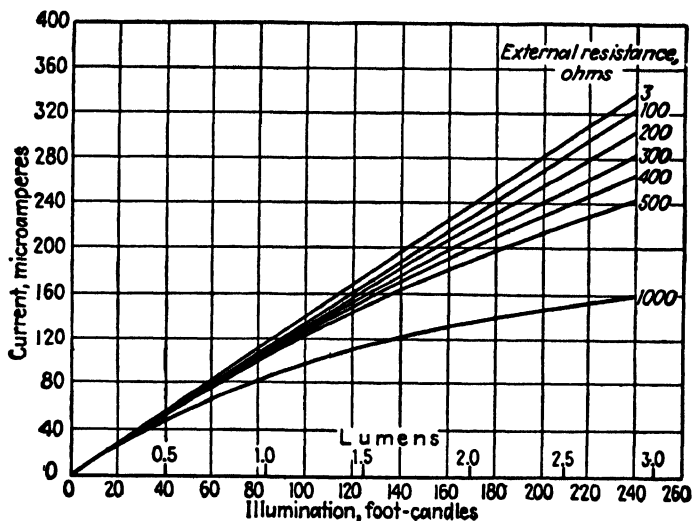



FIG. 124.—Characteristics of a photovoltaic cell, showing the effect of external resistance on the output current.

When light shines through the conducting layer into the light-sensitive material, electrons are forced out of the light-sensitive material and into the iron disk. This makes the light-sensitive material and the conducting layer positive, and makes the iron disk negative. Thus if an external circuit is provided, when light falls on the photocell, conventional current will flow through the external circuit from the conducting layer to the iron disk. If the external resistance is low, the current flow is almost directly proportional to the illumination intensity, as Fig. 124 illustrates. The photocell will supply a much larger current than can be obtained with the simple phototube. It is possible to build sensitive relays and instruments that will operate directly from the output of photovoltaic cells without batteries, tubes, or other equipment.

The simple light meter used in photography is an example of the use of a photocell.

SUMMARY

In this chapter vacuum tubes were considered from the electronic viewpoint, with stress placed on what happens inside the tube. The term "vacuum tube" includes both high-vacuum tubes and gas tubes. The four ways of liberating electrons for use in vacuum tubes are (*a*) by thermionic emission, (*b*) by secondary emission, (*c*) by field emission, and (*d*) by photoelectric emission.

Thermionic cathodes are heated to a high temperature and the electrons are "boiled out" from the metal. The cathodes are (*a*) the directly heated filament type and (*b*) the indirectly heated type, using a separate heater. The directly heated types may be filaments of pure tungsten for use in large transmitting tubes, filaments of thoriated tungsten for use in tubes of intermediate size, or oxide-coated filaments for small tubes. Indirectly heated cathodes are oxide-coated.

When a cathode is heated to operating temperature, electrons in large numbers are thrown off into the space surrounding the cathode. These form into a relatively thin sheath called the "negative space charge." When the anode, or plate, of a diode is made positive, electrons are drawn from the space charge to the positive plate. These electrons then flow on around the external circuit and back to the cathode.

The grid of the triode is between the plate and cathode and is very effective in determining the plate-current flow. If the grid is positive, it increases the current flow. If the grid is negative, it reduces the current flow. The grid is the control electrode. For may purposes the grid is biased or made negative with respect to the cathode. For amplification, the alternating signal voltage is impressed in series with the source of bias voltage. The signal voltage then makes the grid alternately less negative and more negative, and this allows more and then less plate current to flow. The plate current then contains two components, (*a*) an average, or direct-current, component, and (*b*) the alternating signal component.

For the purposes of circuit design, it is necessary to know the constants or coefficients of a vacuum tube. These coefficients are (*a*) the amplification factor μ, (*b*) the plate resistance r_p, and (*c*) the mutual conductance g_m. These coefficients are related as follows:

$$\mu = r_p g_m, \qquad g_m = \frac{\mu}{r_p}, \qquad \text{and} \qquad r_p = \frac{\mu}{g_m}.$$

In a tetrode a second grid, called a "screen grid," is added between the control grid and plate. This grid reduces the capacitance between the plate and control grid and adapts the vacuum tube to use at radio frequencies. The screen grid usually is made positive and accelerates the electrons. When these strike the plate, secondary emission occurs; and if the plate is at a potential less than the screen grid, the secondary electrons from the plate may flow to the screen grid. This gives the tetrode undesired characteristics, and for this reason the plate voltage is kept always above the screen-grid voltage. Because the screen grid shields the plate, the plate is ineffective in controlling

the electron current flow, and the tetrode has a high amplification factor and a high plate resistance.

A third grid, called a "suppressor grid," is added between the screen grid and plate in a pentode. This grid is tied to the cathode, and it does not prevent secondary emission from the plate; but it does prevent the secondary electrons from flowing back to the screen grid. For this reason, in a pentode the plate voltage may fall to a low value, and it need not be kept above that of the screen grid. In the pentode, the plate is rendered even less effective by the presence of the suppressor grid, and the tube has a very high amplification factor and plate resistance.

Remote-cutoff, or variable-mu, tubes are produced by using an unsymmetrical grid arrangement. The coefficients of such tubes vary widely and they can be used in automatic volume-control circuits.

Tubes containing more electrodes than a triode are called "multielectrode," or "multigrid," tubes. They are of two types: (*a*) tubes, such as the tetrode and pentode, which perform functions which cannot be performed by a triode; and (*b*) tubes, such as a diode-triode combination, which are in reality two tube structures in a single envelope.

Vacuum tubes can be classified on the basis of the purpose for which they are to be used. Thus there are the small voltage-amplifying tubes and the large power-output tubes. Triodes may be used for either purpose. In general, tetrodes are used for voltage amplification. Pentodes are used for both purposes. A special tube known as the "beam-power" tube has the characteristics of a power pentode but uses no suppressor grid.

Gas diodes contain an inert gas, such as argon or mercury vapor. The electrons flowing to the anode strike the neutral atoms and cause ionization by collision. The massive positive ions accumulate near the cathode and neutralize the space charge. As a result, the voltage drop across the tube is lower than for the high-vacuum type. Such tubes are used as rectifiers.

In the gas triode (Thyratron), a grid gives control over the start of current flow. Once current starts, however, the grid becomes surrounded by a positive space charge, and the grid loses control. In the gas tetrode, a second grid makes it possible to shift the characteristics of the tube.

Cold-cathode tubes exhibit a constant voltage drop over a wide current range. Such tubes are used for voltage regulators. In the cold-cathode three-electrode tube, a starting anode can be used to initiate the discharge in the main gap, and the tube operates as a relay. The electrons are probably pulled out of the cold cathode by the action of the positive ions that accumulate near the cathode surface. Also, positive ions striking the cold cathode probably liberate electrons.

The phototube consists of a light-sensitive cathode and a positive anode either in a vacuum or an inert gas. The light striking the cathode liberates electrons, which then are attracted to the positive anode. The number of electrons liberated is directly proportional to the light intensity. In the gas phototube, ionization by collision adds to the current flow, producing a more sensitive phototube. A secondary emission multiplier is incorporated in one type of phototube. The phototube acts like a light-controlled resistance.

The photovoltaic cell generates its own voltage when light shines on it, and

no external source of electric energy need be used as with the phototube. A photovoltaic cell is capable of directly operating a sensitive relay.

REVIEW QUESTIONS

1. Explain the difference between a high-vacuum tube and a gas tube.
2. Enumerate and briefly explain four ways in which electrons are liberated from metals.
3. What are the two general types of vacuum tubes?
4. What metals are used for filament cathodes?
5. What is meant by an oxide coating, what are its advantages, and on what types of cathodes is it used?
6. Briefly discuss the type of cathode that would be used in a large high-vacuum tube.
7. Briefly discuss the type of cathode that would be used in a large mercury-vapor tube.
8. What is the direction of electron-current flow inside a high-vacuum triode with a negative grid and a positive plate? What is the direction of conventional-current flow?
9. Discuss the use of the grid in the triode, and compare its effectiveness with that of the plate in affecting the electrons in the space-charge region.
10. What is meant by the terms "amplification factor," "plate resistance," and "mutual conductance"?
11. What is meant by grid bias?
12. What are the advantages and disadvantages of the screen grid in a tetrode?
13. Explain how the suppressor grid functions in the pentode. Why is the suppressor grid usually tied to the cathode?
14. What is meant by a remote-cutoff tube? By what other names is it known?
15. What tubes are classified as multigrid, or multielectrode, tubes?
16. Briefly explain the difference between a voltage-amplifying tube and a power-output tube. Can power be drawn from a voltage-amplifying tube?
17. How does the beam-power tube operate?
18. Explain the function of the gas (or vapor) in a gas diode.
19. Why does the grid lose control in a gas triode?
20. Why are two grids sometimes used in gas tubes?
21. How are electrons drawn from a cold cathode?
22. Explain how the cold-cathode three-electrode gas tube operates as a relay.
23. Briefly compare the principles of operation of high-vacuum and gas phototubes.
24. How does the multiplier phototube operate?
25. How do the phototube and the photovoltaic cell differ in basic principles of operation?

PROBLEMS

1. A small receiving tube has an oxide-coated filament cathode that draws 0.06 ampere at 2.0 volts. What is the power input? A large power-trans-

mitting tube has a tungsten filament that draws 225 amperes at 27 volts. What is the power input? Calculate the power input to the indirectly heated cathode discussed on page 203.

2. The oxide-coated length of the cylinder of an indirectly heated cathode of a small voltage-amplifying high-vacuum tube is 2.0 centimeters and the outside diameter is 0.115 centimeter. A power input of 1.0 watt per square centimeter of coated surface will emit electrons at the rate of 10.0 milliamperes per square centimeter of coated surface. If the filament is operated with this rate of power input, what power in watts will it draw, and what will be the total electron current output in milliamperes.

3. Use Fig. 101*a* and the 180-volt and 150-volt curves to calculate the amplification factor at 1.0, 5.0, and 10.0 milliamperes.

4. Use Fig. 101*a* and the 150-volt curve to calculate the mutual conductance at 1.0, 5.0, and 10.0 milliamperes.

5. Use Fig. 101*b* and the −16-volt curve to calculate the plate resistance at points corresponding to P, P_1, and P_2.

6. For a small triode, the manufacturer gives the following data: At operating plate voltage of 90 volts and grid voltage of −4.5 volts, $\mu = 9.3$ and $r_p = 11{,}000$ ohms. At 90 volts and −9.0 volts, $\mu = 9.3$ and $r_p = 10{,}300$ ohms. Calculate the corresponding values of g_m.

7. Obtain values from tube manufacturers' manuals, and make a table showing μ, r_p, and g_m for typical voltage-amplifying triodes, tetrodes, and pentodes, and for typical power-amplifying triodes and pentodes. Use tubes of comparable physical size and power-handling capacity.

8. Refer to the data given on page 203, and calculate the *total* power loss within the tube during operation.

9. Using the data plotted in Fig. 116, calculate the corresponding values of resistance between points 2–3, and plot these calculated values with current on the X axis and resistance on the Y axis.

10. Assume that the multiplier-type phototube of Fig. 123 is operated so that each primary electron produces S secondary electrons when it strikes each multiplier anode n. What will be the multiplication of the tube? Numerically, what would be a typical value of the total multiplication for a tube of this type?

CHAPTER VII

RECTIFIERS

In radio, it often is necessary to change alternating current into direct current. An important example of this is in the operation of vacuum tubes. As explained in the preceding chapter, direct voltages usually are impressed on the plate and the grids. To avoid the use of batteries, these direct voltages are obtained by changing alternating-current power from commercial power systems to direct-current power.

A **rectifier** is a device that converts alternating current into direct current. The internal characteristics of a rectifier are such that current readily flows through it in one direction, but little or no current flows through it in the other direction. That is, if a direct voltage is impressed on a rectifier, current readily flows for one polarity, but negligible current flows if the voltage connections are reversed. Or if an alternating voltage is impressed on a rectifier, current flows for one half cycle, but negligible current flows for the other half cycle.

A **rectifier unit** is the entire rectifying apparatus, including transformers, filters, and necessary auxiliary apparatus. This is the standard definition (see footnote, page 33). However, the term "rectifier" often is used to designate the entire assembly of equipment.

The Ideal Rectifier.—For most purposes the ideal rectifier is a device that will pass current in one direction without opposition, but will pass no current in the other direction. When voltage of one polarity is applied, such a rectifier will offer zero resistance, and when the polarity of the applied voltage is reversed, the rectifier will offer infinite resistance.

The characteristics of an ideal rectifier are shown in Fig. 125. Also, the graphic symbol is included; the direction of low resistance and large current flow is as indicated by the arrow. For simplicity, this may be designated the *forward*, or *positive, direction*. Thus for the positive direction of applied voltage, the current through a rectifier will rise toward infinity if the resistance of the circuit

approaches zero. For the negative direction of applied voltage, little, if any, current will flow.

When an alternating voltage is applied to a rectifier, current flows for the positive half cycle, but little or no current flows for the negative half cycle. This is illustrated by Fig. 126. Three cycles of the applied voltage are shown, and the rectifier permits only the positive half cycles to cause current flow through the current-limiting resistor. In a sense, the rectifier is a switch which opens the circuit on the negative half cycle, that is, on the half cycle in which no current flows.

FIG. 125.—Characteristics and symbol for an ideal rectifier.

It is apparent that the wave form of the applied voltage is different from the wave form of the voltage ($e = iR$) which appears across the resistor. That is, a pure sine-wave voltage has been applied on the input side, and a half-wave voltage appears on the output side. (In this sense, resistor R is regarded as the device to which the rectified current is being supplied.) A

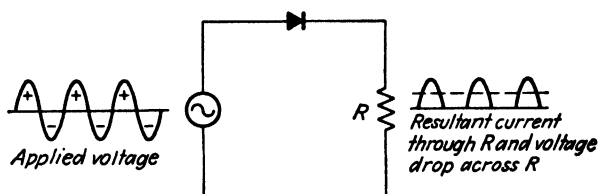
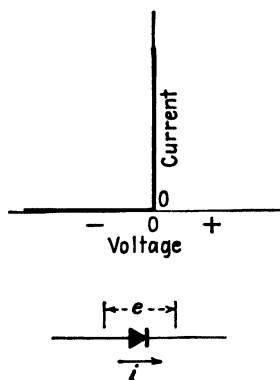



FIG. 126.—When an alternating voltage is impressed on a rectifier, current flows for the positive half cycles, but not for the negative half cycles. The dotted line indicates approximately the average value or rectified direct current.

rectifier may be regarded as a *distorter*, because **distortion** is defined as a change in wave form (see footnote, page 33).

Distortion often is regarded as something always to be avoided, but this is not true. As defined, distortion is a change in wave form, and it may be very much desired, as in the case of rectification. There are many other instances in radio where distortion is desired, and of course there are many instances where the wave form is to be maintained and distortion is undesired. This dual viewpoint regarding distortion is important, and it should be acquired by students of radio.

Basic Rectifier Circuits.—There are many types of rectifier circuits, particularly if rectification is regarded in the broad sense that includes other applications than merely the conversion of alternating to direct current for power-supply purposes. One of the basic circuits was considered in the preceding section and is shown again in Fig. 127.

The Half-wave Rectifier.—The input voltage is increased or decreased as desired by the input transformer. For instance, if the rectifier unit is to charge low-voltage automobile storage batteries, the transformer that is selected would step down the voltage from, say, 115 volts to a few volts; however, if the rectifier unit is to supply plate voltage for vacuum tubes, the transformer that is selected would step up the voltage from 115 volts to several hundred volts.

FIG. 127.—A half-wave rectifier. The equation for the sinusoidal voltage impressed on the transformer is $e = E'_m \sin 2\pi ft$. The rectified voltage or current contains many components. For the half-wave rectifier these are

(*a*) A direct-current component $(E_m/\pi) = 0.318\ E_m$.

(*b*) A fundamental component $(E_m/2) \sin 2\pi ft = 0.5 \sin 2\pi ft$.

(*c*) A second harmonic $(2\ E_m/3\ \pi) \cos 2\ \pi 2ft = 0.21 \cos 2\ \pi 2ft$.

(*d*) A forth harmonic $(2\ E_m/15\ \pi) \cos 2\ \pi 4ft = 0.042 \cos 2\ \pi 4ft$.

(*e*) Additional harmonics of the series of negligible importance.

The voltage E_m is the voltage E'_m stepped up or down as desired by the transformer. The voltage E_m is the total voltage induced in series in the transformer secondary.

As explained in the preceding section, current will flow for one half cycle, and not for the other, producing a **half-wave current** flow through the resistor and a **half-wave voltage** across the resistor, as indicated in Fig. 127.

Here it is well to insert a word of warning: *The wave shape of the current and voltage output of a rectifier sometimes is influenced by the type of load* (whether resistive, inductive, or capacitive), and *the discussions now being given apply to rectifiers working into loads offering pure resistance* (such as R of Fig. 127).

In Fig. 127, the impressed input voltage is a pure sine wave that contains only one frequency term. In contrast, the rectified output current and voltage are nonsinusoidal waves that contain the following: (*a*) a direct-current component E_m/π; (*b*) an alternating-current component $(E_m/2) \sin 2\pi ft$, of magnitude $E_m/2$ and frequency f, the same as that of the wave being rectified; (*c*) an alternating-current component $(2E_m/3\pi) \cos 2\pi 2ft$, of magnitude

$2E_m/3\pi$ and frequency $2f$, twice that of the wave being rectified; and (*d*) an alternating-current component $(2E_m/15\pi) \cos 2\pi 4ft$, of magnitude $2E_m/15\pi$ and frequency $4f$, four times that of the wave

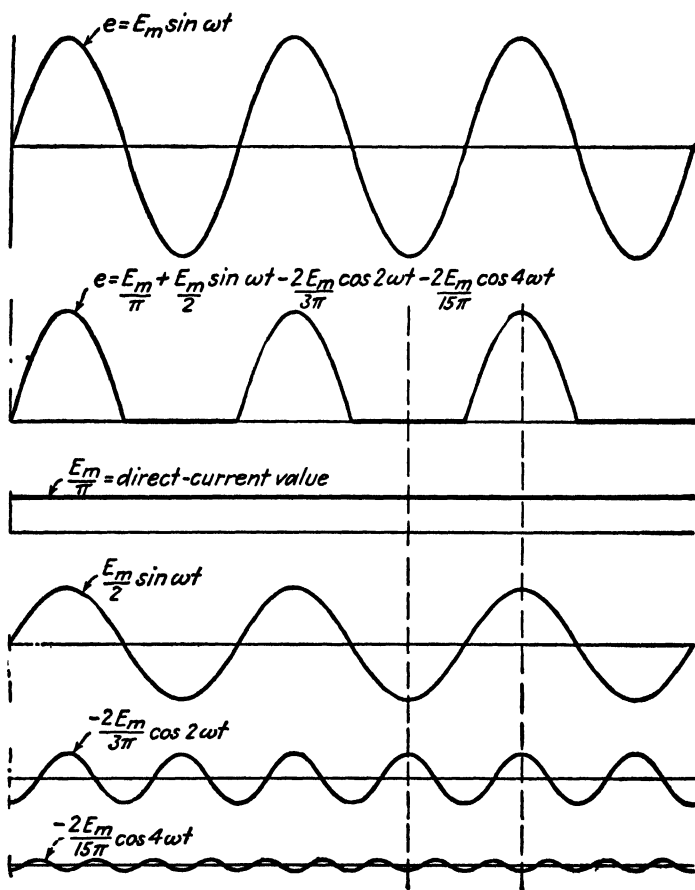



FIG. 128.—When the sine-wave voltage shown at the top is impressed on a half-wave rectifier, the wave shape of the voltage across a resistance load will be as shown by the second curve. This half-wave voltage will contain the components shown by the third, fourth, fifth, and sixth curves. This can be proved by adding corresponding instantaneous values. These curves also represent the current through the load resistor of Fig. 127.

being rectified. Of course higher frequency components also exist, but usually these are negligible. In all these components, E_m is the maximum value of the alternating voltage that is impressed on the rectifier. An ideal rectifier is assumed.

To prove that these components actually do exist in a rectified

half wave, Fig. 128 has been included. If at any point along the X axis the various instantaneous positive and negative values of the components are combined, the rectified half wave results. The question arises: Where do the various components come from? They are not present in the impressed input. The answer is that the various components are created in the process of distortion, caused by the rectifier which allows only one half cycle of the impressed voltage wave to force current through the circuit. The rectifier is a distorter, and such a distorter produces components not present in the input. *This is a very important principle* and is of much use in communication. If a telephone receiver, or a loud-speaker, is connected across R of Fig. 127, hum will result, indicating the presence of the various components, or harmonics as they are called. In this circuit (*a*) a direct-current term, (*b*) a term of the same frequency as the impressed voltage, (*c*) a second harmonic, and (*d*) a fourth harmonic all exist in the output (together with other harmonics that have been neglected). The frequency and magnitude of each alternating component can be determined with a **wave analyzer**, which is an instrument which can be tuned to detect the presence of each harmonic in a nonsinusoidal wave. This wave analyzer is similar to a radio-receiving set (page 565), which can pick out a given radio-signal frequency from the many radio signals which exist simultaneously in space.

The Full-wave Rectifier.—The basic circuit of the full-wave rectifier is shown in Fig. 129. The transformer increases, or decreases, the voltage impressed on the rectifier so that the direct current and voltage outputs are of the magnitudes desired.

At the instant the upper end of the transformer secondary is positive, current will flow around the upper loop and through the resistance. At this instant the lower end of the transformer secondary is negative, and no current flows around the lower loop. On the next half cycle of the applied voltage the lower end of the transformer secondary is positive and the upper end is negative. During this time interval, current flows around the lower loop and through the load resistor as indicated, but no current flows through the upper loop. As a result, *each* half cycle of the applied voltage forces current *in the same direction* through the load resistor, and a rectified **full-wave current** flows through the resistor. The voltage across the resistor also will have the same shape (since $e = iR$), as Fig. 129 indicates.

The impressed voltage is a sinusoidal voltage that contains only

one frequency term. The rectified output current and voltage are nonsinusoidal waves that contain many components. These are listed on Fig. 129. As for the half-wave rectified current, or voltage, the first part of each component is the magnitude, and the second part gives the frequency. Thus for the full-wave rectifier working into a resistance load, the components are (*a*) a direct-current component of magnitude $2E_m/\pi$; (*b*) a second harmonic of frequency $2f$ and of magnitude $4E_m/3\pi$; (*c*) a fourth harmonic of

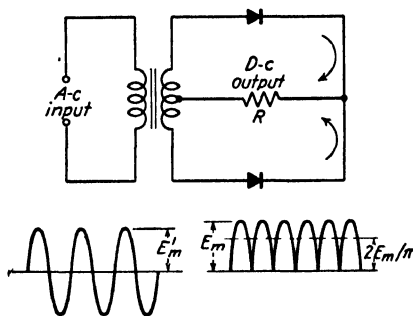



FIG. 129.—A full-wave rectifier. The equation for the sinusoidal voltage impressed on the transformer is $e = E'_m \sin 2\pi ft$. The rectified voltage or current contains many components which for the full-wave rectifier are
(*a*) A direct-current component $(2E_m/\pi) = 0.637 E_m$.
(*b*) A second harmonic $(4E_m/3\pi) \cos 2\pi 2ft = 0.425 E_m \cos 2\pi 2ft$.
(*c*) A fourth harmonic $(4E_m/15\pi) \cos 2\pi 4ft = 0.085 E_m \cos 2\pi 2ft$.
(*d*) A sixth harmonic $(4E_m/35\pi) \cos 2\pi 6ft = 0.036 E_m \cos 2\pi 6ft$.
(*e*) Additional harmonics of the series of negligible importance.
The voltage E_m is voltage E'_m stepped up or down by the transformer. The voltage E_m is the voltage induced in series in *one half* of the transformer secondary.

frequency $4f$ and magnitude $4E_m/15\pi$; and (*d*) a sixth harmonic of frequency $6f$ and magnitude $4E_m/35\pi$. Higher frequency components also exist, but they usually are negligible. In all these components E_m is the maximum value of the alternating voltage impressed on *each rectifier when it is conducting;* thus E_m is one half of the transformer secondary voltage. Ideal rectifier units are assumed.

Rectifiers.—As defined on page 219, the rectifier is the device that converts alternating current into direct current. This section will consider such devices.

The Ideal Rectifier.—This would have zero resistance in the conducting direction and infinite resistance in the nonconducting direction. The characteristics and the equivalent circuit were shown in Fig. 125.

The Crystal Rectifier.—Certain types of crystals have rectifying properties, examples being silicon crystal rectifiers and germanium crystal rectifiers, although many other substances may be used. In a typical crystal rectifier, the crystal is held in a housing, and a fine tungsten wire point makes firm contact on the crystal surface. Good design and assembly are of importance to ensure a stable contact, a typical arrangement being shown in Fig. 130*a*. The characteristics of a typical crystal detector are as shown in Fig. 130*b*; the characteristics of a specific assembly depend on the nature of the crystal, the contact, the pressure, etc. As noted, some current flows in the "nonconducting" direction in most crystals, and sometimes this reverse-current flow is a considerable percentage of the forward-current flow, that is, the flow in the conducting direction. An idealized characteristic curve for a perfect crystal is shown in Fig. 130*c*, and the circuit equivalent in Fig. 130*d*. The resistance shown is necessary to give the positive voltage–current characteristics of the crystal; that is, the crystal has appreciable internal resistance. Crystal rectifiers are used extensively at very high radio frequencies.

FIG. 130.—Cross section and characteristics of a typical crystal rectifier.

The Barrier-layer Rectifier.—If cuprous oxide is formed properly on the surface of a copper electrode, often in the form of a disk, the combination has rectifying properties and is called a **copper oxide rectifier.** The conventional direction of current flow is from the oxide to the copper. A somewhat similar rectifier uses an iron-selenium combination and is called a **selenium rectifier.** Another rectifier uses a copper sulphide-magnesium combination, and is called a **copper sulphide rectifier.** Rectification is

thought to occur at a barrier layer that exists between the different substances. The electrodes may be disks or of any convenient shape. Each rectifying element will stand but a few volts, and stacks of rectifying elements in series are used for rectifying large voltages. Parallel stacks are used for giving high current capacities. The characteristics of a copper oxide rectifier unit are shown in Fig. 131. It should be remembered that these are for one specific rectifier, that they are built in a variety of sizes and capacities, and that they have different resistance characteristics. Rectifiers of this general type have low resistance in one direction and high resistance in the other direction. They therefore fall into the same general classification as the rectifier of Fig. 130.

An iron-selenium rectifier rated at 600 milliamperes and 100 volts for the complete unit, or 600 milliamperes and 5 volts per disk The units of this rectifier are interconnected for a bridge rectifier circuit This is *not* the rectifier whose characteristics are shown in Fig 131. (*Federal Telephone and Radio Corporation, manufacturing subsidiary of International Telephone and Telegraph Corporation*)

The High-vacuum Diode.—This was discussed in detail in the preceding chapter. Electrons flow to the plate when it is positive with respect to the cathode, but will not flow for the opposite polarity. Conventional current flow is, of course, in the reverse direction. There is a voltage drop between the cathode and plate, or anode, and thus the characteristics, curves, and equivalent circuit of the high-vacuum diode are as shown in Fig. 130.

FIG 131 —Characteristics of a copper-oxide rectifier of a type used for modulation and demodulation (Chap. XI) in carrier-telephone apparatus.

The Gas Diode.—This tube was discussed in the preceding chapter, and its characteristics were given in Fig. 112, page 202. Plotting these curves with the

axes changed gives Fig. 132*a* as the actual characteristics, Fig. 132*b* as the idealized characteristics, and Fig. 132*c* as the equivalent circuit. The battery indicated is of course a fictitious source of voltage. It is necessary in the equivalent circuit because the cur-

A high-vacuum diode having the following ratings. filament voltage, 21.5 volts; filament current, 41 amperes; maximum peak inverse voltage, 25,000 volts; and maximum peak plate current, 5 amperes. This tube has a water-cooled anode or plate shown in the lower half of the photograph. Such tubes are used as high-voltage rectifiers for the plate supply of large power tubes in radio transmitters. (*Western Electric Co.*)

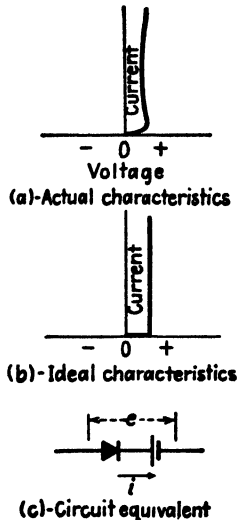
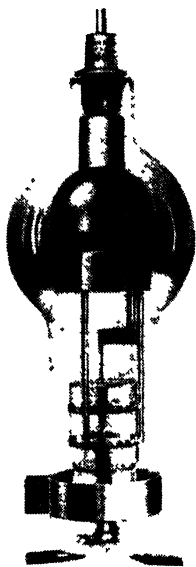



Fig. 132.—Characteristics of a gas diode rectifier.

rent flow is "offset" as compared with Fig. 130, and appreciable current does not flow until the anode is positive by an amount equal to the opposing effect of the battery.

Of course there are many other types of rectifiers, but those discussed are the types most widely used.

Inverse Voltage.—It is quite evident that in the selection of a crystal, or a barrier-layer rectifier, or a diode, consideration would be given to the current-carrying capacity of the device. The

current-carrying capacity would, of course, be largely determined by the heat-radiating ability, that is, by its physical size and the amount of air circulation, etc. What is not so apparent, however, is the fact that much consideration must be given to the amount of voltage that a rectifier will stand without failure.

A gas diode having the following ratings: filament voltage, 5 volts; filament current, 42 amperes; approximate anode-cathode drop, 15 volts; maximum peak plate current, 20 amperes; and maximum peak inverse voltage, 20,000 volts. This tube is cooled by radiation and normal air circulation. Such tubes are used as high-voltage rectifiers for the plate supply of large power tubes in radio transmitters. (*Western Electric Co.*)

Thus in the preceding chapter it was shown that the voltage drop across a small high-vacuum diode rectifier tube was perhaps 50 volts *when the tube was conducting* and that the voltage drop across any gas tube was only about 15 volts *when the tube was conducting.* The question that is important, yet not so apparent, is this: What is the voltage across a rectifier tube, or other rectifier, when the tube is not conducting?

To answer this question, reference is made to Fig. 127. When the negative half cycle occurs and the tube does not conduct, there is no current flow in the circuit, no voltage drop across the load resistor, and all of the transformer secondary voltage is impressed in the inverse (or nonconducting) direction on the rectifier. If a condenser is connected across the resistor of Fig. 127, and if the resistor is of high resistance, the voltage impressed across the tube on the negative half cycle will be almost equal to *twice* the transformer secondary voltage. This is because the condenser charges on the positive half cycle and retains its charge during the negative half cycle and because the polarities of the condenser and of the transformer voltage on the negative half cycle are such that the voltages add.

For a full-wave rectifier, Fig. 129 should be examined. When the upper rectifier is conducting, the lower one is not, and vice versa. The idle rectifier and the operating rectifier are effectively *in series* across the *entire* secondary voltage. When conducting, the voltage drop across a rectifier

is low (if it were not, it would be a poor rectifier for most purposes). The inverse voltage across the idle rectifier is the transformer secondary voltage minus the drop in voltage across the operating rectifier, and this is essentially the *entire secondary voltage.*

Now the important point is this: During the nonconducting half cycle the voltage is tending to cause current to flow through a rectifier in the nonconducting, or inverse, direction. For either the half-wave or full-wave rectifying circuits, this inverse voltage is at least the *entire* secondary voltage. Any rectifier must be capable of withstanding without failure the peak or maximum value of the inverse voltage. Thus it is common in rating rectifiers to specify the **peak inverse anode voltage** that the device will stand with safety.

Effect of Load on Rectifier Output.—A very important application of a rectifier is to convert alternating-current power to direct-current power in order to supply some device such as a vacuum-tube amplifier. Usually a filter is connected between the rectifier and the load, the purpose being to pass to the load the direct-current component of the rectified wave and to attenuate all alternating-current components and to prevent their reaching the load.

The nature of the input impedance of the load, and of the filter with connected load, greatly affects the operation of the rectifier. The performance of rectifiers with typical loads will be considered now.

Resistance Loads.—Rectifiers with these loads were illustrated in Figs. 126, 127, and 129. The full-wave rectifier is more widely used than the half-wave type, and will, accordingly, be discussed. As indicated in Fig. 129, and as explained in the accompanying discussions, the current through the load resistor and the voltage across the resistor consist of a direct-current component and various alternating-current components, or harmonics. The magnitude and frequency of each component is as given there.

Now when a current flows through a resistor, it is considered that a voltage forces it through the resistor. Thus if a rectifier causes a current containing many components to flow through a resistor, it may be considered that a rectifier is a device consisting of several voltage sources in series, that each of these forces its respective current component through the resistor, and that the summation of these various components gives the rectified current (Fig. 128). Thus the equivalent circuit for a rectifier with a pure resistance

load (such as Fig. 129) is shown in Fig. 133. This discussion neglects the impedance of the transformer secondary and the internal impedance of the rectifier. For a pure resistance load, each voltage component of the rectified voltage wave forces a current component of corresponding magnitude through a load of pure resistance. There is no reduction in the harmonic content of the current, that is, no filtering action with pure resistance.

Resistance Load with Series Inductance.—As previously stated, the purpose of a rectifier is to convert alternating-current power to direct-current power. In doing this, the input wave is distorted and, in addition to the desired direct-current component, alternating-current components, or harmonics, exist in the rectifier output. With a load of pure resistance (Fig. 133), these harmonic components flow through the load. Generally, it is desired to allow only the direct-current component to flow through the load and to reduce the flow of harmonics to negligible values.

If an inductor L of high inductance and low resistance is inserted in series with the load resistor, as shown in Fig. 134, the direct-current component will flow through the resistor without appreciable opposition by the coil, but the flow of harmonics will be drastically opposed by the inductive reactance of the coil. As a result, the magnitude of each alternating-current component through the resistor will be much less than if the inductor L were not in the circuit. The current through the resistor and the voltage across the resistor then will be free to a large extent from alternating-current components. In practice it is commonly stated that the effect of an inductor is to "smooth out" the wave so that it contains but "little ripple." If the alternating-current components are not filtered out, hum will result in the device, such as a radio set, that is being energized by the power supply.

Illustrative Problem.—The effective value of the entire secondary voltage of the transformer of Fig. 129 is 750 volts, 60 cycles. Assume that the transformer secondary impedance is negligible and that the voltage drop across the rectifier is negligible. Calculate the magnitude of each current and voltage component delivered to a 10,000-ohm resistor when the resistor is used alone, and when it is in series with a 12-henry inductor, or "choke," that has 200 ohms direct-current resistance. Also calculate the percentage ripple in each case.

Solution.—Step 1. Calculate the magnitude of each current and voltage component and of the percentage ripple, when the resistor is used alone, remembering that for a full-wave transformer only one-half of the entire transformer secondary voltage appears across each rectifier and load

when conduction occurs, and that the equations are written in terms of maximum, not effective values. Thus, for the full-wave rectifier of Fig. 129, $E_m = (750/2) \times 1.414 = 530$ volts. From the explanation accom-

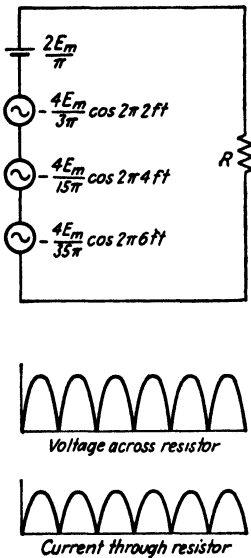



FIG. 133.—A rectifier across a resistance load acts like a direct voltage and several (actually an infinite number of) alternating harmonic voltages in series. The negative signs are necessary to give the components the proper phase relations. A similar diagram applies to the half-wave rectifier and resistance load, but the various components are as given by Fig. 127.

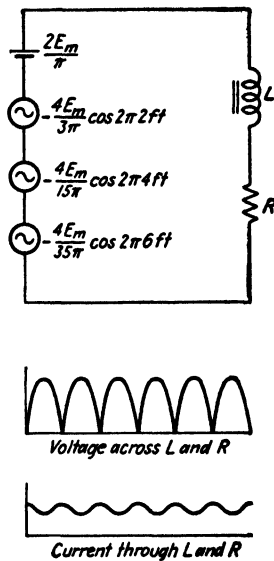



FIG. 134.—A rectifier across a load composed of resistance R and inductance L in series acts like a direct voltage and several (actually an infinite number of) alternating voltages in series. The negative signs are necessary to give the proper phase relations. A similar diagram applies to the half-wave rectifier, but the components will be as given in Fig. 127.

panying Fig. 129, and from the discussion in the text accompanying Fig. 133, the components are as follow:

Direct current: $I = 0.637 \times \dfrac{530}{10{,}000} = 33.7$ milliamperes;

$E = 0.0337 \times 10{,}000 = 337$ volts.

Second harmonic: $I = 0.425 \times \dfrac{530}{10{,}000} = 22.5$ milliamperes;

$E = 0.0225 \times 10{,}000 = 225$ volts.

Fourth harmonic: $I = 0.085 \times \dfrac{530}{10{,}000} = 4.5$ milliamperes;

$E = 0.0045 \times 10{,}000 = 45$ volts.

Sixth harmonic: $I = 0.036 \times \frac{530}{10{,}000} = 1.9$ milliamperes;
$E = 0.0019 \times 10{,}000 = 19$ volts.

In calculating the percentage ripple it is customary to neglect the effect of all but the first alternating component. The percentage ripple is the ratio of the *effective* value of the alternating component to the value of the direct component *at the load*, the ratio being multiplied by 100 to express it in percentage. If current values are used, the percentage ripple is $(22.5 \times 0.707)/33.7 = 0.47$, and $0.47 \times 100 = 47$ per cent. The factor 0.707 was used because the values given for the harmonics are maximum values since E_m, the maximum voltage, was used in the computations.

Step 2. Calculate the magnitude of each current and voltage component and the percentage ripple when the 12-henry 200-ohm choke coil is in series with the 10,000-ohm resistor, as in Fig. 134. It is necessary first to calculate the impedance offered by the coil to each frequency component. In doing this the resistance is so small that it may be neglected and the impedance may be assumed equal to the reactance.

Impedance for direct-current component equals zero if resistance is neglected.

Reactance for second harmonic $= 2\pi fL = 6.28 \times 2 \times 60 \times 12 = 9050$ ohms.

Reactance for fourth harmonic $= 2\pi fL = 6.28 \times 4 \times 60 \times 12 = 18{,}100$ ohms.

Reactance for sixth harmonic $= 2\pi fL = 6.28 \times 6 \times 60 \times 12 = 27{,}150$ ohms.

It is now necessary to calculate the total impedance offered by the 12-henry choke and the 10,000-ohm load resistor in series.

Total impedance for direct-current component equals 10,000 ohms (approximately).

Total impedance for second harmonic $= \sqrt{(10{,}000)^2 + (9050)^2} = 13{,}500$ ohms.

Total impedance for fourth harmonic $= \sqrt{(10{,}000)^2 + (18{,}100)^2} = 20{,}700$ ohms.

Total impedance for sixth harmonic $= \sqrt{(10{,}000)^2 + (27{,}150)^2} = 28{,}900$ ohms.

The current and voltage components at the load resistance will be:

Direct current: $I = 0.637 \times \frac{530}{10{,}000} = 33.7$ milliamperes;
$E = 0.0337 \times 10{,}000 = 337$ volts.

Second harmonic: $I = 0.425 \times \frac{530}{13{,}500} = 16.7$ milliamperes;
$E = 0.0167 \times 10{,}000 = 167$ volts.

Fourth harmonic: $I = 0.085 \times \frac{530}{20{,}700} = 2.2$ milliamperes;
$E = 0.0022 \times 10{,}000 = 22$ volts.

Sixth harmonic: $I = 0.036 \times \frac{530}{28{,}900} = 0.7$ milliampere;

$E = 0.0007 \times 10{,}000 = 7$ volts.

The percentage ripple is $[(16.7 \times 0.707)/33.7] \times 100 = 35$ per cent.

From these calculations it is clear that a choke connected in series with a load resistor will reduce the alternating-current components that pass through the resistor and will reduce the percentage ripple. For a practical case, however, the use of a single choke as previously indicated would be of little value. In other words, if the "resistor" used to represent the load to which the rectifier was supplying power were in reality a vacuum-tube amplifier, then far more filtering of the rectified wave would be required.

Resistance Load with Parallel Capacitance.—Two cases thus far have been considered; the first was a full-wave rectifier with pure resistance load, and the second was a full-wave rectifier with an inductor or choke in series with the resistance load. It was shown that with these two circuits the current through the load and the voltage across the load could be computed for each component, and also that the percentage ripple could be found. These calculations were made using elementary circuit theory. If a condenser is placed *across* a load resistor, then the calculation of the harmonics and the percentage ripple becomes rather involved. The reasons for this will now be explained.

A full-wave rectifier with a load composed of resistance and parallel capacitance is shown in Fig. 135. This circuit is the same as that of Fig. 129 except that the position of the load R has been rearranged and condenser C has been added. In studying Fig. 135, two extreme conditions will be investigated. (*a*) If it is assumed that capacitor C has a very low value, then its effect is of little consequence, and the current that flows through the load resistor R will be just the same as without the condenser; that is, it will be the same as the rectifier with a pure resistance load. (*b*) If the condenser has a very large value, then its effect is of much importance. A condenser will charge essentially to the peak of the rectified wave, and the large condenser will tend to hold its charge and maintain a constant direct voltage across the load resistor. Thus the condenser would be very effective in "smoothing" out the current and voltage at the resistor, and in reducing the percentage ripple.

As stated in the preceding paragraph, a large condenser would

be very effective in preventing the flow of harmonics through a load resistor. Note that a *parallel* condenser accomplishes this by attempting to hold the voltage across the resistor constant, whereas the *series* inductor previously considered attempted to prevent the flow of the harmonics and thus hold the current constant. The way in which the parallel capacitor attempts to hold the voltage constant is shown in Fig. 135. The circuit is assumed to have been in operation for some time and to have reached the steady-state condition.

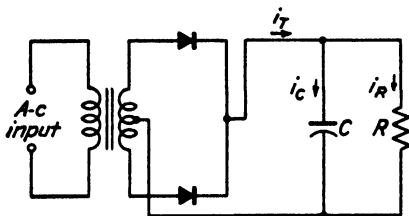
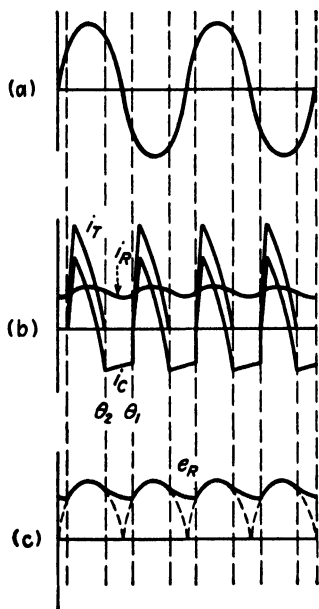



FIG. 135.—When the sinusoidal alternating voltage (*a*) is impressed on the rectifiers, the currents that flow are as shown by (*b*), and the current i_R causes the voltage drop e_R to appear across the parallel circuit composed of the condenser and load resistor. This is the rectified voltage wave.

The rectifier units are receiving an alternating voltage (Fig. 135*a*) from the transformer secondary and are *attempting* to produce a full-wave rectified pulsating voltage across the parallel resistor and capacitor, as indicated by the broken line of Fig. 135*c*. As the voltage rises, the condenser takes a peaked charging current i_C, as shown by Fig. 135*b*. When the voltage reaches the maximum value, the condenser current reaches zero, because current flows through a condenser only when the voltage changes. When the voltage from the rectifier units starts to fall, the condenser discharges. But the condenser cannot discharge *back* through the rectifier units because they conduct in one direction only. Thus the condenser discharges through the resistor. The total current i_T supplied by the rectifier units is the instantaneous sum of the condenser

current i_C and the resistor current i_R. When the point on the impressed voltage cycle is reached where the condenser is supplying all the current required by the resistor, then the current through the rectifier *drops to zero* and **cutout** occurs, as indicated at point θ_2. After the condenser has discharged somewhat, and the impressed voltage again rises, the condenser once more takes a charge, and **cutin** occurs, as at point θ_1. As a result of the attempt of the condenser to maintain the voltage constant across the load resistor, the current taken by the resistor is as indicated by i_R of Fig. 135*b*, and the voltage across the resistor (and condenser in parallel) is $e_R = i_R R$, as shown by curve e_R of Fig. 135*c*.

It is recognized that this action is involved, and further study[1] is recommended for a complete understanding. From the practical standpoint, the important points are the following: (*a*) With a capacitor of sufficient size, current flows through the rectifier in spurts, and these current impulses are quite peaked. This is *very* important when gas tubes are used for rectifiers, because these peaks must not exceed maximum current ratings or the tubes will be ruined. (*b*) The condenser changes the wave form of the voltage impressed on the load resistor. Instead of the voltage resembling the broken curve of Fig. 135*c*, it resembles e_R. A wave such as e_R contains fewer harmonics than the broken curve, and thus fewer harmonics flow in the load resistor. In this way a condenser causes filtering action, and reduces the percentage ripple. (*c*) The shape of the wave e_R determines the harmonic content. The shape is determined by the constants of the circuit. For any combination of resistance and capacitance the wave has a particular shape. For these reasons it is extremely difficult to calculate the magnitudes of the haromonics and the percentage ripple in circuits similar to Fig. 135*c*. Curves and experimental data usually are employed.

Rectifier Filters.—A complete system for converting from alternating-current power to direct-current power involves the units shown in Fig. 136. The purpose of the filter is to suppress the alternating-current components of the rectified wave and to keep these harmonics from flowing through the load. If these harmonics are permitted to flow to the device being energized by the power supply, then hum may result. Yet the filter must permit the

[1] For additional information, consult M. B. Stout, Analysis of Rectifier Filter Circuits, *Electrical Engineering*, September, 1935.

direct-current component to reach the load without serious attenuation. Thus a rectifier filter must have the form of the low-pass filter of Fig. 87, page 165.

The simplest possible forms of filtering circuits are the series inductor and the parallel capacitor of the preceding section. But as was shown for the inductor, and as also is true for the capacitor, ordinary chokes and condensers alone do not produce sufficient filtering for most purposes. However, combinations of coils and condensers in low-pass filter circuits are *very* effective in reducing the ripple of rectified currents and voltages. The filter systems used are known in radio as the "choke-input" type and the "condenser-input" type. A word of warning is necessary here: It is the electrical action, rather than the physical arrangement, that determines

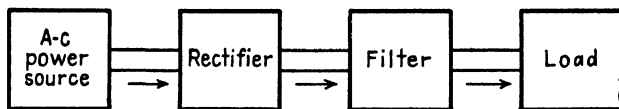



FIG. 136.—Elements of a system of rectification for furnishing a direct voltage and direct current to a load.

whether a filter *is operating* as a choke-input filter or as a condenser-input filter.

Choke-input Filter.—A full-wave diode rectifier arranged with a choke-input filter is shown in Fig. 137. The transformer increases or decreases the alternating supply voltage to give the value of E_m and, hence, the direct-voltage and current output that is desired. Often this is a center-tapped 110- to 750-volt transformer with windings having adequate current-carrying capacity. A low-voltage high-current cathode-heating winding usually is provided. On the half cycle in which the upper end of the transformer is positive, electrons flow from the cathode to this positive plate, and *conventional* current flows in the opposite direction. On the next half cycle the lower plate conducts. This gives the polarity indicated.

The inductors usually are placed in the positive lead because filtering is more effective. Because the coils have low resistance, the desired direct current readily flows through the load resistor R. There are several ways of explaining filter action, and one is this: The series inductors offer high impedance to the flow of harmonics and thus hold the flow of alternating-current components to low

values. The parallel capacitors offer paths of low impedance to current flow and tend to short-circuit, or by-pass, alternating currents. Since $E = IZ$, and for a large condenser Z approaches zero, the alternating-current voltage that can exist across a large condenser is small. Thus it can be said that the parallel capacitors tend to hold the voltage constant across the circuit. By this combined action of coils which tend to hold the current constant and condensers which tend to hold the voltage constant, the percentage ripple can be reduced to negligible amounts. An alternate explanation for rectifier filter action is that the circuit follows low-pass filter theory (page 165) and that the filter has a cutoff frequency

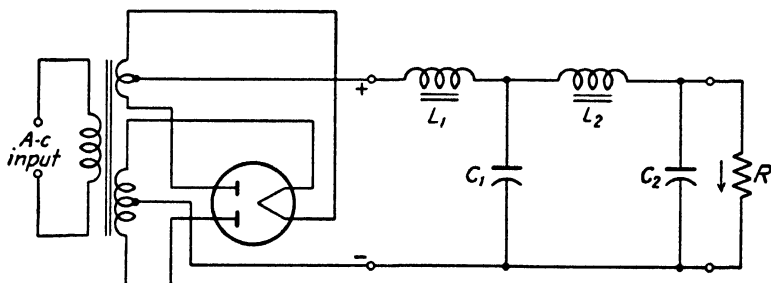



FIG. 137.—A full-wave rectifier using a twin diode tube. The filter is of the choke-input type.

considerably below the lowest frequency component present in the rectified wave.

A filter may be *connected* with choke input, but *may not* be operating as a choke-input filter. Thus, if L_1 of Fig. 137 is of low inductance and C_1 of high capacitance, cutout as explained in the preceding section may occur. Cutout is also affected by other constants of the circuit, which will be considered in the following pages. If cutout does not occur, then the rectified voltage wave across the filter input remains as in Fig. 129, and by simple circuit theory the magnitude of each component that reaches the load resistance can be calculated.

Condenser-input Filter.—If coil L_1 is removed, or if an additional condenser is connected between L_1 and the rectifier tube, then the filter has the form of a condenser-input filter. If the condenser has sufficient capacitance, then the circuit operates much like the rectifier with a load of resistance and capacitance in parallel, which was considered in the preceding section. Of course if the input

condenser has insufficient capacitance then the circuit will not operate as a condenser-input filter. Cutout also is affected by other constants of the circuit, which will be explained in the following section.

If a circuit is operating as a condenser-input filter, then the voltage wave across the filter input is not the simple rectified full wave, as shown in Fig. 129 and by the broken line of Fig. 135*c*. Instead, the voltage wave across the filter input is as shown by e_R of Fig. 135*c*. With choke-input operation it is relatively easy to calculate the magnitude of the direct current and voltage at the load. Also, it is relatively easy to calculate the magnitude of each alternating-current component and the percentage ripple at the load. For a condenser-input filter this is difficult to do because the shape of the rectified voltage wave—and hence the components it contains—varies with the circuit constants. By contrast, if a circuit is operating as choke input, the content of the voltage wave is known.

Preventing Cutout.—It has been mentioned that a circuit may be *connected* as a choke-input filter, but *may operate* as a condenser-input filter in which case cutout occurs. Also, it has been mentioned that a circuit may be *connected* as a condenser-input filter, but *may operate* as a choke-input filter. These statements undoubtedly have been confusing, and the question has undoubtedly arisen: What difference does it make anyway? The answer is that an abrupt change in voltage occurs when conditions change from choke-input operation to condenser-input operation, as will be explained.

If the curves of Fig. 135*c* are examined, it is evident that the voltage wave e_R for condenser-input operation has greater area under it than has the broken curve for choke-input operation. This means that a given rectifier with condenser-input filter will supply a greater direct current and direct voltage to a load than will a comparable rectifier, using the same supply transformer, but with choke-input filter. And furthermore, the input current and input voltage (i_R and e_R of Fig. 135) are "smoothed out" quite effectively by the input condenser, and for this reason a given amount of filtering is more effective because the wave initially is "smoothed out."

Higher direct current and voltage output with the same transformer (or the same output with a transformer of less turns) and

more effective filtering seem to indicate that condenser-input filters should be used always. Nevertheless, there are several objections to the use of these filters: (a) A condenser-input filter draws a very peaked current, as curve i_T of Fig. 135 indicates. This is not serious with high-vacuum tubes, which usually can stand such operation. However, gas tubes cannot stand such operation and soon would be ruined; in general, gas tubes are operated with choke-input filters. If they are not, then care must be taken to ensure that the allowable peak anode current is not exceeded. For instance, a suitable protective impedance may be connected in the anode circuit. (b) If not properly designed, a filter may operate as choke input when the load is connected, and may change over to condenser-input operation if the power is reduced or the load is disconnected. This causes the direct output voltage to rise abruptly on light loads and no loads, and often this is very undesirable. For example, when such a circuit is used with a radio telegraph transmitter (page 474), the load on the rectifier unit fluctuates between wide limits as the sending key is operated. With a circuit that changed over from choke- to condenser-input operation, large load-voltage fluctuations would occur, and these are undesired.

To prevent cutout from occurring with a filter *arranged* with choke input, the size of inductor L_1 and the magnitude of the load resistance R of Fig. 137 are related by the equation

$$L_1 = \frac{R}{1000} \qquad \text{and} \qquad R = 1000\,L_1, \tag{75}$$

where L_1 is in henrys and R is in ohms. This is the practical form used; the theoretical value, derived on the basis that the current through the rectifier elements never falls to zero during the conducting half cycle, is about 1100. Thus a filter *arranged* with choke input will always operate as choke input without cutout and the accompanying voltage rise if the conditions of Eq. (75) apply.

Sometimes a resistor called a **bleeder** is connected permanently across the filter output to prevent cutout irrespective of the resistance value of the actual load drawing the power. Thus if the inductor L_1 of Fig. 137 is 12 henrys, then from Eq. (75) a resistor of $R = 12 \times 1000 = 12{,}000$ ohms is the critical value of load resistance. If it is desired to change R to 25,000 ohms, then the inductance must be $L_1 = 25{,}000/1000 = 25$ henrys or cutout will

occur and the voltage will rise. However, if a bleeder of such value that the bleeder resistance in parallel with the load resistance always satisfies Eq. (75) is connected across the filter output, then cutout cannot occur, and the voltage regulation is good. By **voltage regulation** is meant

$$\frac{\text{No-load voltage} - \text{full-load voltage}}{\text{Full-load voltage}} \times 100. \qquad (76)$$

Thus, if the direct voltage output with no load is 375 volts, and the direct voltage when full load is being drawn is 350 volts, the regulation is $[(375 - 350)/350] \times 100 = 7.15$ per cent. This is rather high, but whether it is too high depends on the particular application.

A resistor, or bleeder, permanently connected across the output of a rectifier filter (that is, across the output of a power supply) serves another and very important purpose. In general, radio power supplies produce dangerously high voltages and considerable energy is stored in the filter capacitors. Many fatalities have resulted among radio workers because they forgot that the filter condensers would retain these dangerous charges for considerable periods of time after the power supply was shut off, the length of time depending largely on the types of condensers used. A bleeder ensures that the condensers discharge immediately when a power supply is shut off, and, if possible, *bleeders always should be installed on power supplies*, and these bleeders should be inspected regularly. Another advantage of the bleeder is that it can be used as a voltage divider, with taps brought out as desired to give various output voltages.

Power-supply Design.—In radio and associated fields, the input transformer, rectifiers, filter, and bleeder are spoken of as the **power supply** (rather than as the rectifier unit, page 219). It should be evident from the preceding pages that the accurate design of a power supply in its entirety would be an exceedingly difficult task, particularly if it were of the condenser-input type in which cutout occurred. On the other hand, if certain approximations are used, then the performance of a power supply can be determined easily and within reasonable limits. Two examples will now be given.

Illustrative Problem.—It is desired to build a power supply that will deliver to a load 200 milliamperes direct current at 400 volts. At times, the load may be very small.

Solution.—Step 1. Select the tube and transformer. It is decided to use a type 83 mercury-vapor diode. This tube will stand a peak inverse voltage of 1550 volts, and will supply a maximum direct output current of 225 milliamperes. It has two anodes in the same glass envelope and can be used for full-wave rectification. The tube drop is about 15 volts. It has a 5-volt 3-ampere filament. The transformer selected is shown in Fig. 138. Since a mercury-vapor tube is to be used, the filter should have choke input. To obtain a direct voltage of 400 volts, the value of E_m (page 224) must be at least $400 = 2E_m/\pi$, and $E_m = 400 \times 3.14/2 = 628$ volts. The effective value of this is $E = 628 \times 0.707 = 445$ volts. To allow for voltage drops in the tube, transformer, and filter chokes, a

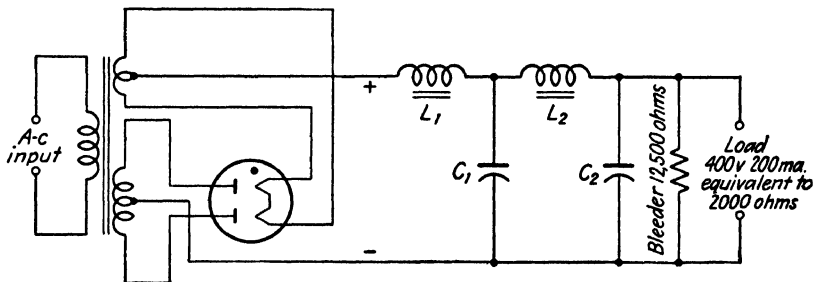



FIG. 138.—A full-wave rectifier using a twin gas diode. A bleeder always should be used with condensers having low leakage to ensure that the condensers discharge immediately after the input power is disconnected.

transformer with a 110-volt primary and a center-tapped 1000-volt secondary is selected from a radio parts catalogue. Of course this transformer must have adequate power-handling capacity, and it should have a 5-volt 3-ampere winding for heating the filament. The peak inverse voltage on the tube will be $1000 \times 1.414 = 1414$ volts, and this does not exceed the tube rating.

Step 2. Determine the inductance of L_1 to prevent cutout and to ensure choke-input operation. The equivalent resistance of a load that draws 200 milliamperes at 400 volts is $R = 400/0.2 = 2000$ ohms. From Eq.(75), $L_1 = 2000/1000 = 2.0$ henrys. This calculation shows that *with full load* a very small amount of inductance will suffice. No-load conditions must be investigated. It is decided to use a 12,500-ohm bleeder. If this is done, from Eq.(75), an input choke of $L_1 = 12{,}500/1000 = 12.5$ henrys must be used. The radio parts catalogue shows that a choke coil having a full-load inductance of 5.0 henrys, a no-load inductance of 12.5 henrys; a full-load current-carrying capacity of 225 milliamperes, and a resistance of 80 ohms is available. (Of course, a different bleeder could be used if the no-load inductance were not 12.5 henrys.)

Step 3. Determine the size of condensers C_1, C_2, and coil L_2. This selection is largely a matter of judgment. The parts catalogue shows that a choke having an inductance of 15 henrys at a current of 225 milliamperes is available, and it is decided to use this for L_2. The catalogue shows that

4-microfarad, 1000-volt, oil-impregnated paper condensers are vailable, and because of past experience it is decided to use these.

Step 4. Calculate the approximate percentage ripple at full load. This usually is calculated for the first alternating-current component, because it is the largest in magnitude, and since its frequency is lowest, the filter is least effective. In other words, if a check shows that the filtering is sufficient at the harmonic of lowest frequency, it usually will be more than sufficient at the other harmonics, except under some special condition where a device might be particularly sensitive to a certain harmonic. Since Fig. 138 is a full-wave rectifier, the first alternating-current component will have a frequency of 120 cycles and a magnitude (Fig. 129) of $4E_m/3\pi$, or for the transformer selected in Step 1, which had a *total* maximum secondary voltage of 1414 volts, $4 \times 707/3 \times 3.1416 = 300$ volts peak, or $300 \times 0.707 = 212$ volts effective value. The reactances of each filter unit should be computed next. These are as follows: $X_{L_1} = 2\pi fL_1 = 6.28 \times 120 \times 5 = 3780$ ohms; $X_{L_2} = 6.28 \times 120 \times 12.5 = 9450$ ohms; and $X_{C_1} = X_{C_2} = 1/(2\pi fC) = 1/(6.28 \times 120 \times 4 \times 10^{-6}) = 332$ ohms. The resistance of the rectifier "load" or filter termination will be the parallel resistance of the 2000 ohms equivalent to the actual load and the 12,500-ohm bleeder, or $R = R_1R_2 \,/\, (R_1 + R_2) = (2000 \times 12{,}500)/(2000 + 12{,}500) = 1725$ ohms.

To find the percentage ripple, the direct voltage at the load and the effective value of the voltage of the harmonic at the load must be known. The direct voltage is to be 400 volts. The harmonic voltage can be found from regular series-parallel circuit theory (page 84) or by the following approximation. As shown in the preceding section, the harmonic voltage across the filter input is 212 volts effective value. Assume that condenser C_1 is short-circuited. The alternating current flow through choke L_1 would be $212/3780 = 0.056$ ampere. Now assume that the short circuit is removed from C_1 and that all this current flows through it. The alternating voltage drop would be $0.056 \times 332 = 18.6$ volts. Next assume that condenser C_2 is short-circuited. Then the current through L_2 will be $18.6/9450 = 0.00197 = 0.002$ ampere, approximately. Next, assume that the short circuit is removed from C_2, and assume that all this current flows through it. Then the voltage drop will be $0.002 \times 332 = 0.66$ volt. On the basis of the above assumptions, this is the effective value of the harmonic voltage across the load. The percentage ripple will be $(0.66/400) \times 100 = 0.165$ per cent. This solution is a rough approximation, but often is of sufficient accuracy. Whether or not 0.165 percentage ripple is sufficiently low depends entirely on the device to which the rectified power is being supplied. A careful study of this solution will show that the tube and the chokes are operated slightly above the rated values, but such apparatus will stand small overloads of this nature.

Two points should be brought out at this time. One regards the nature of the inductors, or chokes L_1 and L_2. Because at full load little inductance is needed in the input choke to prevent cutout, the no-load inductance and full-load inductance, or so-called **incremental inductance,** of the choke may vary widely. This will be

discussed in the next section. The second point in that considerable radio-frequency radiation may occur with gas tubes used in rectifiers. Thus if gas or vapor tubes are used in the vicinity of sensitive radio-receiving sets, or are used to supply the tubes of radio-receiving sets, then the tubes must be shielded and radio-frequency chokes must be placed in the supply leads.

Illustrative Problem.—It is desired to build a power supply that will deliver up to 100 milliamperes direct current at 300 volts. The regulation is not important.

Solution.—Step 1. Curves are available in the tube manuals published by the manufacturers. These manuals give data that are useful in the selection of the tube and the transformer. From such a manual it is decided to use a 5U4-G, a high-vacuum diode with two plates permitting full-wave rectification. The curves for this tube are reproduced in Fig. 139. Since a high-vacuum tube rather than a gas tube is to be used, the filter will have condenser input. From Fig. 139 it is seen that this tube will give 100 milliamperes at 325 volts at the filter input with 300 volts effective (rms) value impressed on the plates. A radio-parts catalogue lists a transformer with a 110-volt primary, a center-tapped 610-volt secondary, and a 5-volt 3-ampere secondary; this is available, and is selected. The 610-volt secondary will give 305 volts for each plate, slightly exceeding the value needed, and will deliver 125 milliamperes.

FIG. 139.—Operation characteristics of a 5U4-G, a high-vacuum twin-diode rectifier tube. (*Data from RCA Receiving Tube Manual* RC 14.)

Step 2. Next, the filter must be "designed." For a power supply of this type experience rather than theory generally is used Thus one experienced in such matters knows that is common to use, say, 8-microfarad electrolytic condensers which have a direct working voltage of 450 volts and a choke which has an inductance of about 12 henrys. The radio-parts catalogue lists such a choke at 125 milliamperes maximum current and 231 ohms resistance. Since the maximum current is to be 100 milliamperes, the direct-voltage drop in the choke will be 23 volts. With the tube and transformer selected, the input to the filter is 325 volts, and hence the voltage at the load will be about 300 volts at 100 milliamperes as desired. The final rectifier will be as shown in Fig. 140.

This method of "design" seems very superficial, and it is, but it represents a practical solution to most problems of this type. A filter of this type is sufficient for most applications. Sometimes

the inductor used is the field coil of the loudspeaker (page 35) in the radio set to which the rectified power is being supplied. No bleeder is used in this instance because electrolytic condensers have rather high leakage, and would soon discharge themselves. In this respect they are unlike oil-impregnated paper condensers, which hold their charges for long periods. Of course a bleeder could be used if desired, and could be tapped at various points and used as a voltage divider.

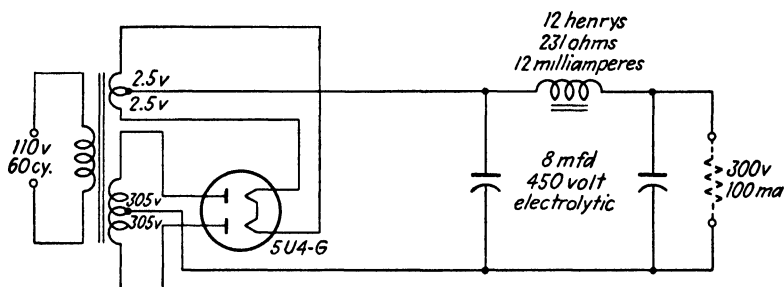



FIG. 140.—A simple full-wave rectifier for delivering a maximum of 100 milliamperes at 300 volts.

Sometimes it is desired that a power supply have a continuously variable output voltage. A convenient circuit for this purpose is shown in Fig. 141. An inductor, or **autotransformer,** that is continuously adjustable is used as an inductive voltage divider across the source. Then the plate voltage impressed on the tubes can be varied continuously between zero and the maximum value. A *separate* filament-supply transformer that steps down from 110 to, say, 5 volts must be used, because the filament voltage must remain constant. A variable resistor can be used to control the input, but a variable inductor (or autotransformer) gives better efficiency and is in general to be desired. A popular device for this purpose is known by the trade name Variac.

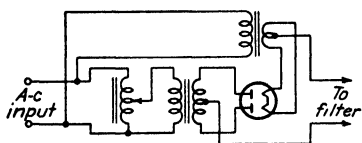



FIG. 141.—Rectifier portion of a power supply providing a continuously variable output.

Incremental Inductance.—This has been mentioned on page 63. It is of much importance in filter chokes that use iron-cored coils, such as the kind employed in power supplies. Such coils carry both direct- and alternating-current components. Of interest is the

self-inductance and the resulting reactance offered in opposition to the flow of the alternating-current components. This reactance opposes alternating-current flow by developing an opposing voltage when the current flows through it. This opposing voltage reduces the current flow to low values. Voltages are caused by changes in magnetic flux. The magnetic flux change caused by current variations depends on the point of operation on the saturation curve of the magnetic material. Where both direct- and alternating-

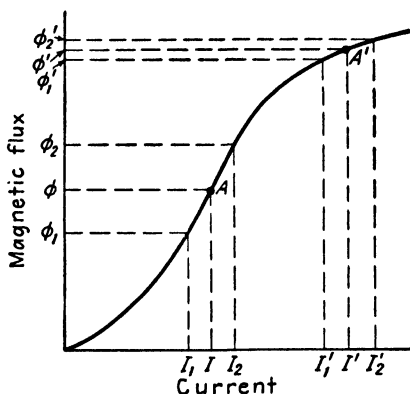



FIG. 142.—If the direct current through an iron-cored coil is I, the point of operation on the magnetization curve will be at A, and an alternating-current component I_1-I_2 superimposed on I will cause a magnetic flux change ϕ_1-ϕ_2. This will give the coil a certain incremental inductance. If, however, the direct current is I', then the point of operation will be A', and an alternating-current component I'_1-I_2 of the same magnitude as I_1-I_2 will cause a smaller flux change ϕ'_1-ϕ'_2, and the incremental inductance will be less.

current components flow through a coil, the point of operation is determined by the direct-current value.

These principles are illustrated in Fig. 142. If the direct-current value is represented by I, and the alternating current, which is simultaneously flowing, is represented by I_1-I_2, then the corresponding flux change will be ϕ_1-ϕ_2, a given back voltage will be induced, and the coil would be said to have a certain value of **incremental inductance.** This name is used because it is the inductance offered to an incremental current change, that is, to a current variation I_1-I_2 about the average value I. If now, the direct-current component is I' and the alternating component is I'_1-I'_2 the same in magnitude as before, then the corresponding flux change will be less than with direct-current I, and the induced voltage and incremental inductance also will be less. In this con-

nection it is well to remember that inductance may be defined in terms of flux linkage per unit current.

An example of a coil of this type is the coil discussed in Step 2 on page 241. This coil had an inductance of 12.5 henrys at no load, and 5.0 henrys at full load of 225 milliamperes. Such an inductor is sometimes called a **swinging choke.** It can be used as the input choke (L_1, Fig. 137) because in order to prevent cutout less inductance is necessary at high load than at low load. The second choke of Fig. 137 often is called the **smoothing choke.** Usually, it is designed so that its inductance is more constant at various values of direct current. A coil with a small core that saturates at full direct current is a swinging choke. A coil with an air gap in the core, or with a very generous core, does not saturate so easily, and, hence, its inductance is more constant. The terms "swinging choke" and "smoothing choke" are misleading, because both chokes contribute to the filtering and smoothing of the rectified wave.

Rectifier Voltage Dividers.—A bleeder across the output of a rectifier filter serves at least two very useful functions: (*a*) it prevents operation from changing from choke input to condenser input, and it prevents a rise in voltage as the load is changed; (*b*) the bleeder ensures that the filter condensers discharge immediately when the device is shut off, thus eliminating a hazard to operating personnel. From the standpoint of preventing cutout, the resistance used should not exceed $R = 1000L_1$ in accordance with Eq. (75). From the standpoint of preventing injury, the bleeder resistance may be any reasonable value. Of course, the bleeder wastes power, and this may or may not be of importance.

As was mentioned on page 240, a bleeder also may be used as a voltage divider (page 61). This makes available various output voltages. The design of a voltage divider that will be used to prevent cutout, discharge the filter condensers, and provide several voltages will be considered now.

Illustrative Problem.—It is desired to arrange the bleeder of Fig. 138 as a voltage divider so that 45 volts at 20 milliamperes, 90 volts at 20 milliamperes, and 180 volts at 50 milliamperes are available. The arrangement is shown in Fig. 143.

Solution.—Step 1. In referring to the discussion accompanying Fig. 138, it is found that the maximum current that can be carried by the equipment is about 225 milliamperes and that the resistance of the bleeder should not exceed about 12,500 ohms if cutout is to be prevented. If

cutout occurs, the voltage will rise and the regulation will be poor. The bleeder can be less than this value, but from the standpoint of reducing the power wasted in it, the bleeder should be as large as possible, not exceeding 12,500 ohms, of course. If the voltage and currents specified are to be supplied, then the bleeder must be less than 12,500 ohms. It must be designed step by step.

Step 2. Purely as an assumption, it is decided that there always will be 32 milliamperes through the lower section of the voltage divider. Thus to give 45 volts across it, the resistance of this section should be $R = E/I = 45/0.032 = 1405$ ohms.

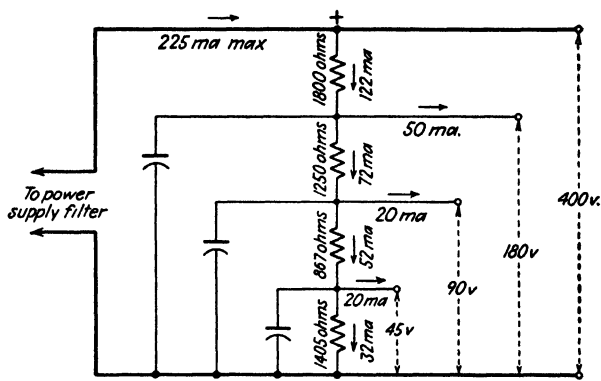



FIG. 143.—Illustrating the design of a voltage divider for a power supply.

Step 3. The current through the second section (from the bottom) will be 52 milliamperes and there must be 45 volts across it. Thus the resistance of this section should be $R = E/I = 45/0.052 = 867$ ohms.

Step 4. The current through the third section will be 72 milliamperes, and 90 volts drop must occur across it. The required resistance is, therefore, $R = 90/0.072 = 1250$ ohms.

Step 5. There will be $400 - 180 = 220$ volts drop across the upper section, and 122 milliamperes through it. The required resistance is $R = 220/0.122 = 1800$ ohms.

These calculations are based on the condition that each tap on the voltage divider (bleeder) is drawing the current at the voltage specified. If any one of these values is changed, then all the others are changed. Of course the power supply cannot deliver 400 volts at 200 milliamperes (page 240) at the same time that it is delivering the currents to the various taps. Also, in the original calculations a few volts extra were allowed to provide voltage drops in the tubes, transformers, etc. In other words, the calculations just made are approximate.

The voltage divider could be built up of four separate resistors

having the values computed. The power-handling capacity of each resistor can be found from the voltage across it and the current through it. In practice, *one* resistor having a total resistance of $1800 + 1250 + 867 + 1405 = 5322$ ohms (or as near that value as was commerically available) would probably be used. Resistors wound on ceramic forms and with continuously adjustable taps are available. The power-handling capacity of the upper end would need to be greater than that of the lower end, because the upper end carries more current. One uniform resistor would be used in most cases. If the 122 milliamperes flowed through the *entire* resistor, the power-handling capacity would have to be $P = EI = 400 \times 0.122 = 32.8$ watts, so a 50-watt adjustable resistor probably would be selected. The final adjustments to determine the positions of the taps would probably be made with milliammeters and a voltmeter, using the current values actually to be drawn.

It will be noted in Fig. 143 that condensers are connected across each resistor. These are to "by-pass" alternating-current components. Another way of looking at the matter is this: If the condensers were so large that they had zero reactance, then since $E = IZ$, no alternating voltage could exist across a condenser *or across the resistor with which it is in parallel.* If this is true, then a change in the current of the load connected to one tap on the voltage divider could not affect the voltage across other taps. To be more specific, suppose that some device connected from − to +90 draws a current having an alternating component. If the condensers were not there, then the voltage from − to +90 would vary slightly at the alternating-current frequency, and this would affect the voltage across the − to +45 tap, because the resistance of 1405 ohms is *common* to both the +45 and the +90 volt tap. With the condensers across these **common impedances**, a current change in one circuit cannot affect the others. Common impedances, not by-passed with condensers, may cause undesired transfer of signal from one circuit, or portion of a circuit, to another. This may cause oscillations in amplifiers, for example. The sizes and types of the condensers used depend on the circumstances, such as the types of circuits and the nature of the signals involved. For ordinary power supplies, several microfarads or more might be used in each capacitor. Experimentation might be necessary to determine the correct value.

Resistance-capacitance Filters.—On page 234, the case of a rectifier supplying power to a resistor shunted by a capacitor was considered. If the resistance is high and consequently draws little current, the capacitor will lose but little charge between charging intervals and will maintain essentially constant voltage across the resistor. Accordingly the condenser causes "smoothing" action. For lower values of resistance the condenser still is effective because the voltage across the resistor load will have the shape of curve e_R of Fig. 135c. Here the condenser has smoothing action because it produces a voltage across the resistor that contains fewer harmonics than the ordinary rectified wave (broken curve of Fig. 135c). For very low values of resistance and large current flow from a rectifier, reasonable sizes of capacitance would have negligible smoothing action, because they could not control the voltage across the resistor and cause it to resemble e_R of Fig. 135c.

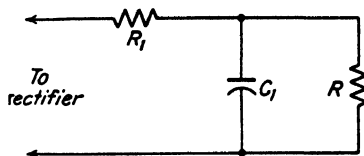



Fig. 144.—Resistor R_1 and capacitor C_1 constitute a resistance-capacitance filter between the rectifier and the load R.

If, however, resistance could be added in series so that the circuit has the form of Fig. 144, quite effective filtering and smoothing would result. In such a circuit cutout would not occur for the values of R_1, C_1, and R usually employed. To calculate the percentage ripple at the load resistance R, the procedure outlined on page 242 could be followed. The magnitude of the harmonics would be as given on page 231 for a full-wave rectifier and no cutout. Condenser C_1 would be assumed to have zero reactance, and the approximate alternating current flowing would be computed. For the second harmonic this would be $I_m = (4E_m/3\pi)/R_1$. Then it would be assumed that C_1 did have reactance, and the harmonic voltage across it (and the resistor) would be $E_m = I X_{C_1}$. From this value (multiplied by 0.707) and the direct voltage across the resistor, the percentage ripple could be found. Additional resistors and condensers, corresponding to L_2 and C_2 of Fig. 138, could be added if necessary.

This resistance-capacitance combination makes an excellent filter for certain applications. It could be made just as effective as the choke-condenser type except for the fact that the choke coil offers high impedance to the alternating harmonic components,

but the choke offers little, and usually negligible resistance, to the direct-current components. For this reason resistance-capacitance filters usually are limited to applications where the direct current flow is low, and where the voltage drop across the series resistors can be tolerated; in other words, to circuits where the voltage delivered to the filter is sufficiently greater than the voltage that must be delivered by the filter to the load.

Voltage Stabilizers.—For some purposes, particularly in measurements, it is desired that the rectified direct-current power supplied to a device remain constant, irrespective of changes in the

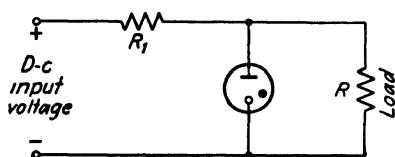



FIG. 145.—A voltage stabilizer composed of a cold-cathode gas diode and resistor R_1.

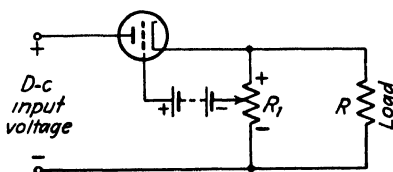



FIG. 146.—A voltage stabilizer using a high-vacuum triode.

alternating power voltage and changes in the magnitude of the load. Sometimes only one of these requirements is of importance. Voltage stabilizers or voltage regulators are used to hold voltages constant.

Voltage Stabilizers Using Gas Tubes.—A simple circuit is shown in Fig. 145. The tube used can be a simple cold-cathode gas diode, such as a neon lamp, or a cold-cathode gas triode. In any event, the tube must have constant voltage-current characteristics, such as those shown in Figs. 116 and 117 of the preceding chapter.

Over their normal current operating range, these tubes *maintain a constant voltage;* thus they will hold the voltage across the load R constant at this value. If the line voltage increases, the current through the gas tube increases, and the voltage rise is absorbed by an added IR drop in resistor R_1. If the line voltage decreases, then the gas tube takes less current, and there is less IR drop in resistor R_1. Also, if the load is changed and the load current increases, the tube current decreases, and vice versa. The circuit of Fig. 145 compensates for changes in supply voltage or load requirements, within the range of the tube. Two or more tubes can be used in series, and tubes with voltage ratings other than those shown in Figs. 116 and 117 are available commercially.

Voltage Stabilizers Using High-vacuum Tubes.—There are several types of these circuits, one being shown in simplified form in Fig. 146. The slider is adjusted until the voltage on the upper portion of the high-resistance voltage divider R_1 overcomes the grid-battery voltage and the grid is negative. The tube then acts like a variable resistance with the plate-to-cathode voltage drop in series between the voltage source and the load resistance R across which the voltage is to be held constant.

If for any reason the voltage across the load R rises, then the voltage across the divider R_1 will be greater, the grid will be made more negative with respect to the cathode, the current passed by the triode will be reduced, and the current taken by the load and the voltage across the load will return to the original value. The reverse action occurs if the voltage across the load falls. This circuit will maintain essentially constant voltage at the load irrespective of changes in the load resistance or of changes in the input voltage (within the limits of operation). A constant-voltage gas regulator tube may be used instead of the battery.

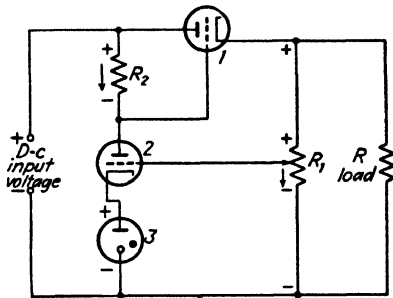



FIG. 147.—A voltage stabilizer of a type extensively used.

A very satisfactory regulator is shown in Fig. 147. Tube 1 is a high-vacuum power-output tube, tube 2 is a high-vacuum voltage-amplifying tube, and tube 3 is a cold-cathode gas-regulator tube. This last tube maintains a constant voltage between the cathode of tube 2 and the negative side of the circuit. The voltage of R_1 is adjusted so that an increase in voltage across load R changes the bias on the grid of tube 2 and causes tube 2 to draw more current through resistor R_2. This increases the bias on tube 1, and decreases the plate current and plate voltage at the load resistor R. A decrease in the output voltage delivered to the load R causes the opposite effect. In a sense, the constant grid-battery voltage of Fig. 146 is a reference voltage, and in the same sense the constant voltage drop across the voltage-regulator tube 3 of Fig. 147 is a reference voltage.

Ballast Tubes or Current Regulators.—These tubes or lamps are not vacuum tubes in the sense that current flows through a high

vacuum or a gas between electrodes. These tubes contain resistances that maintain a *current* constant. Because of this action they can be used as current regulators.

There are several types of ballast tubes available. A popular type consists of an iron wire in a glass bulb filled with hydrogen. As the current flow through the iron wire is increased from zero, the temperature is raised, and soon a portion of the wire glows a dull red. At this temperature, the surrounding hydrogen gas absorbs the heat and conducts it away from the dull red portion of the wire and to the walls of the glass bulb as fast as the heat is

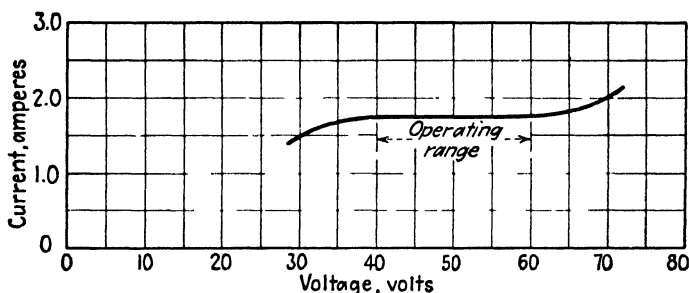



FIG. 148.—Characteristics of a typical ballast tube used for holding a current constant.

generated at that point. Hydrogen, among the usable gases, is an excellent conductor of heat.

As the voltage is increased, more of the iron wire heats a dull red, and this increases the resistance and keeps the current from increasing. If the voltage is increased again, more of the wire reaches a dull red, and again the increased resistance maintains the current constant. As a result of this action, a ballast tube has the characteristics of Fig. 148. Note that with this tube, over the operating range an increase in voltage *does not* cause an increase in current. The ballast tube is a *constant-current device.* This is in contrast with the gas-regulator tube, which is a constant-voltage device.

The gas-regulator tube is placed *across* circuits to maintain the voltage constant. The ballast tube is placed *in series* with circuits to maintain constant current to the circuit. Ballast tubes of the proper size are placed in series with the cathodes of tubes to maintain constant current flow. A ballast tube placed in series with the primary of a transformer will maintain the current constant

and compensate for relatively slow line-voltage variations. Ballast tubes are sluggish in action compared with vacuum tubes because they are temperature-operated devices.

Voltage-doubling Rectifiers.—For most purposes, vacuum tubes of the ordinary radio-receiver type require plate potentials of several hundred volts. If the usual 110- to 120-volt 60-cycle power is rectified, the direct voltage is insufficient for many tube applications. This means that in many instances the power-line voltage must be increased before rectification so that the direct-voltage output ($2E_m/\pi$ for the full-wave rectifier) will be sufficiently large. The voltage applied to the rectifiers can be increased by a transformer. But, transformers are costly, heavy, occupy space, and in certain small radio receivers it is desired to build rectifiers without transformers.

Fig. 149.—The voltage-doubling rectifier. This rectifier gives a direct-voltage output about twice the maximum value of the impressed alternating voltage. No transformer is required.

The so-called "voltage-doubling rectifier" will produce voltages of several hundred volts with ordinary 110 to 120 volts, 60 cycles, applied, and *without* the use of a transformer. The actual circuit, equivalent circuit, and output voltage of a typical voltage doubler are shown in Fig. 149. The operation can be explained from the equivalent circuit as follows: During that part of the alternating-voltage cycle when terminal 1 is positive and terminal 2 is negative, *conventional* current will pass through the upper rectifier and charge the upper condenser, making the upper plate positive. (It may be helpful here to imagine that a resistor of high resistance is connected directly across each condenser.) On the negative half cycle when terminal 1 is negative and terminal 2 is positive, the lower rectifier will conduct, and the lower condenser will be charged with its upper terminal positive.

By the above action, during *one cycle* each condenser has been charged in the same direction to essentially the peak value of the alternating voltage wave. Since these two condensers are in

series insofar as the output terminals are concerned, the output voltage will be approximately twice the peak value of the impressed voltage wave. That is, a 60-cycle 110-volt supply will give a direct voltage of about 110 × 2 × 1.414 = 310 volts. Of course this assumes ideal rectifiers and condensers, and it assumes that no power is drawn.

If a resistor or other similar load is connected across the output terminals, the condensers do not remain charged to the peak value of the alternating voltage, but current from the condensers flows through the load and the voltage varies as Fig. 149 indicates. The

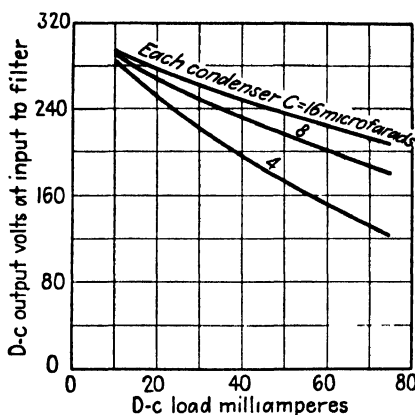
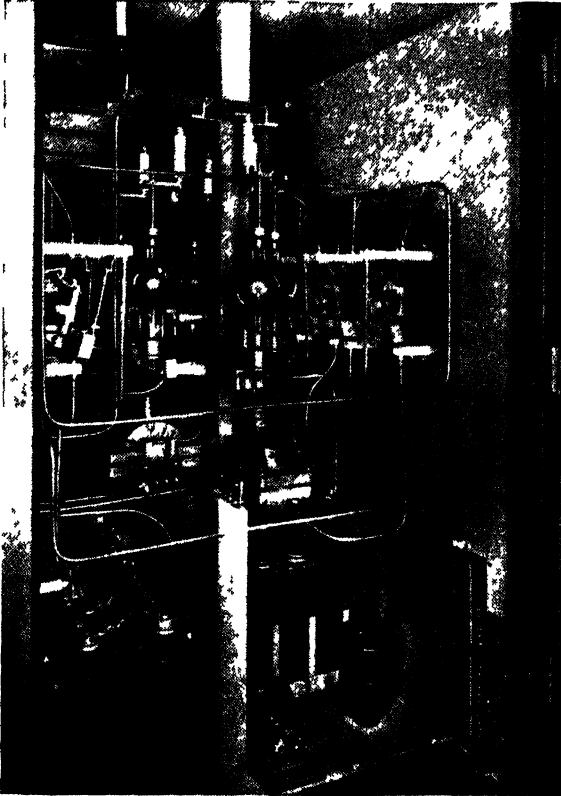



FIG. 150.—Output of a 25Z6 as a voltage-doubling rectifier. This is a high-vacuum twin diode. (*Data from RCA Receiving Tube Manual* RC 14.)

capacitance of the condensers will influence the output voltage available at various current values, as indicated in Fig. 150. A suitable filter can be connected between the voltage doubler of Fig. 149 and the load. If the voltage and current needed by the load are not great, a resistance-capacitance filter may be used. Voltage doublers are used in many applications outside radio; for example, in operating X-ray apparatus. Voltage triplers and voltage quadruplers also are possible, but are in limited use.

Polyphase Rectifiers.—Thus far the rectifiers that have been considered obtained their alternating-current power from single-phase 60-cycle sources. These rectifiers were for the purpose of furnishing relatively small amounts of direct-current power, such as that required by radio-receiving sets, audio-frequency amplifiers, and small radio transmitters requiring a maximum of a few hundred watts. The rectifiers that deliver direct-current power of the order

of a kilowatt and above to large radio transmitters usually draw alternating-current power from three-phase power systems. These are known as **polyphase rectifiers.** There are many types of polyphase rectifiers, but attention will be confined to two basic types.

A polyphase rectifier of the type used in a large radio-broadcast transmitter. This is a six-phase rectifier, employing gas tubes. A spare tube is always maintained with a heated cathode and can be substituted instantly for any tube that becomes faulty. Filter condensers and other apparatus also are shown. (*Westinghouse Electric Corporation.*)

Three-phase Rectifiers.—The connections of a common three-phase rectifier are shown in Fig. 151. For simplicity, only the secondary windings of the three-phase power transformer are shown. The primary windings that are connected to the three-phase 60-cycle power source are of course magnetically connected with the secondary windings by being placed on a common laminated silicon-steel core.

Now in the full-wave single-phase rectifier system extensively studied in the preceding pages, the voltages impressed on the two rectifier tubes are 180° out of phase; that is, when the anode of one tube is made positive, the anode of the other is made negative. The result is that one tube passes current, and then the other tube passes current. Because of this, the first alternating-current component, or harmonic, in the rectified output wave has twice the frequency of the alternating-current power that is being rectified.

In a three-phase system, the voltages impressed on the tubes are not 180° out of phase, but are 120° out of phase with the voltage

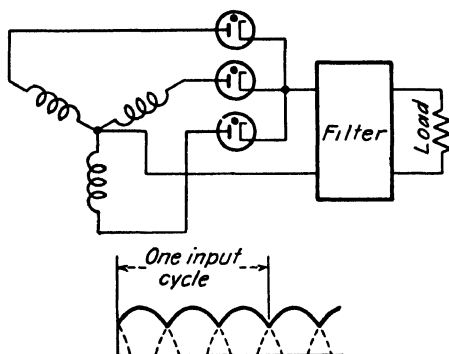



Fig. 151.—A three-phase rectifier. The output from the tubes is as shown by the heavy line in the lower figure.

on the anode of the preceding tube. As a result, one tube is carrying considerable current when the other tube starts, and instead of the rectified current and voltage resembling the waves shown in Figs. 127 or 129, the rectified wave is as shown by the heavy curve in Fig. 151.

It is apparent that the output of the polyphase rectifier is "smoother" than that of a half-wave rectifier or of a full-wave rectifier, and that the output wave more nearly resembles that of the constant direct current and voltage desired. In fact, the situation is this: In the rectified output of the half-wave rectifier the lowest frequency component has the same frequency as the alternating wave being rectified; in the full-wave rectifier, the lowest frequency component is twice that of the wave being rectified; and in the three-phase rectifier, the lowest frequency component is *three times* the frequency of the wave being rectified. The higher the frequency, the greater is the reactance of the series

filter chokes, and the smaller the reactance of the parallel capacitors. Hence, given values of inductance and capacitance are more effective in filtering the rectified wave from a three-phase rectifier than they are in filtering half-wave and full-wave outputs. A three-phase system has other advantages also, such as its general suitability for handling large amounts of power.

Six-phase Rectifiers.—The simplified circuit of a six-phase rectifier is shown in Fig. 152. Here again the primary windings have

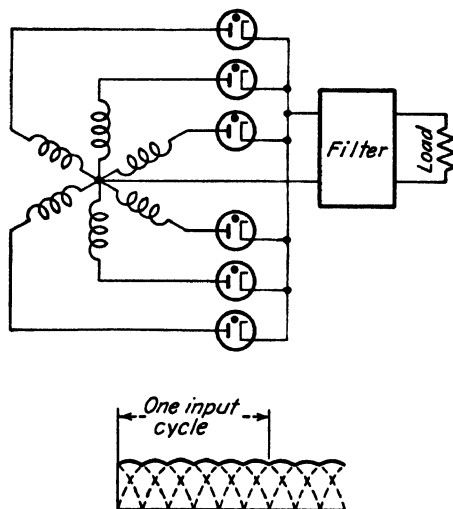



FIG. 152.—A six-phase rectifier. The output from the tubes is as shown by the heavy line in the lower figure.

been omitted. In this circuit the voltages applied to each tube are 60° out of phase with that applied to the preceding tube. As a result, at each instant three tubes are carrying current as Fig. 152 indicates. The rectified output wave is very "smooth," as indicated by the heavy curve of Fig. 152. The lowest harmonic (and thus the most difficult one to filter out) is six times the frequency of the alternating-current supply, further simplifying the filtering.

Polyphase rectifiers have many other advantages, but an advanced knowledge of polyphase power systems is necessary to understand most of them. Suffice it to say, again, that polyphase systems are more suitable for handling large amounts of power than are single-phase systems.

The tubes shown in Figs. 151 and 152 have a black dot in them, indicating that they are gas diodes. Of course high-vacuum diodes could be used. In general, mercury-vapor diodes are used for this purpose because they are more efficient than high-vacuum types. Where large amounts of power are taken, efficiency is of importance. For example, a typical radio-broadcast transmitter that delivers 5 kilowatts to the antenna draws about 18 kilowatts from the 60-cycle supply system. Of course the 13 kilowatts lost is in the entire system, not in the rectifiers alone. The point to be made is that even a moderately sized transmitter draws a reasonably large amount of power, and rectifier efficiency is of importance.

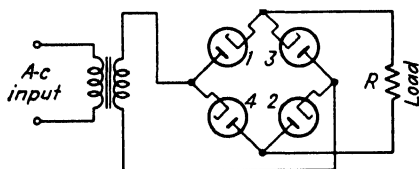



FIG. 153.—A bridge rectifier. The transformer increases or decreases the voltage as desired. The transformer may be omitted, yet full-wave rectification results.

Where small amounts of power are involved, efficiency may be of little or no importance because the cost is negligible.

Bridge Rectifiers.—Vacuum tubes and other rectifiers often are used in bridge circuits to rectify alternating currents. Such circuits are used extensively in measuring instruments where copper oxide rectifiers are used in the bridge arms.

The rectifier unit of Fig. 153 is an example of a bridge rectifier using vacuum tubes, which may be of either the high-vacuum or gas type. The transformer is to increase or decrease the voltage to give the rectified direct voltage desired. The transformer may be omitted, and the bridge circuit connected directly to the alternating-current source.

To explain the operation, assume that the upper end of the transformer secondary is positive and that the lower end is negative; then, *conventional current* (not electron current) will flow up through tube 1, over to the right and down through the resistor *R*, representing the load, back to the left and up through tube 2, and hence back to the source. On the negative half cycle when the upper end of the transformer secondary is negative, and the lower end is positive, conventional current will flow up through rectifier 3, down through the load, up through rectifier 4, and back to the

source. In this way, the bridge rectifier of Fig. 153 produces full-wave rectification.

Plate-voltage Supply from Storage Batteries.—In the preceding pages, methods of obtaining direct voltages from alternating-power systems have been considered. The rectifiers discussed can be designed to supply direct voltages for the plates of vacuum tubes. In radio receivers these direct voltages seldom exceed several hundred volts, but for large radio transmitters they are several thousand volts.

Radio transmitters and receivers very often are operated in remote locations where commercial 60-cycle power service is not available. Also, transmitters and receivers often are installed in automobiles, airplanes, and boats where 60-cycle power may not be available. Of course it is possible and entirely satisfactory for many purposes to use dry batteries for supplying the power, but in some instances this is not the best method.

In automobiles, airplanes, and in other installations it is common practice to operate receivers and transmitters from the storage-battery system. Also, in remote installations where gasoline is available it is common practice to operate radio equipment from storage batteries charged by gasoline-motor electric-generator sets.

If a direct-current motor designed to be driven by storage batteries is mechanically connected to a small 60-cycle alternating-current generator of suitable power capacity, then a radio receiver or a radio transmitter designed to receive power from a commercial 60-cycle power source can be operated. Such methods sometimes are used, but are inefficient because the direct-current power would be changed to alternating-current power by the motor-generator set, and then back to direct-current power by the rectifier associated with the transmitter or receiver.

A more suitable method is to design the radio transmitter or receiver (or amplifier in the case of a sound-truck installation) for operation from direct-current power sources. In this instance, the cathodes of the tubes would be heated directly by the battery, and the plates would be energized by one of the two following ways.

Motor-generator Sets.—If the power required were fairly large, such as that needed to power a small transmitter or a large radio receiver, then the direct voltage for the plates would probably be obtained from a motor-generator set. This would consist of a low

voltage direct-current motor, powered by the storage batteries, and connected to a direct-current generator that gave the high voltage needed for the plates of the tubes. In certain devices the motor and generator are combined into one unit called a **dynamotor.**

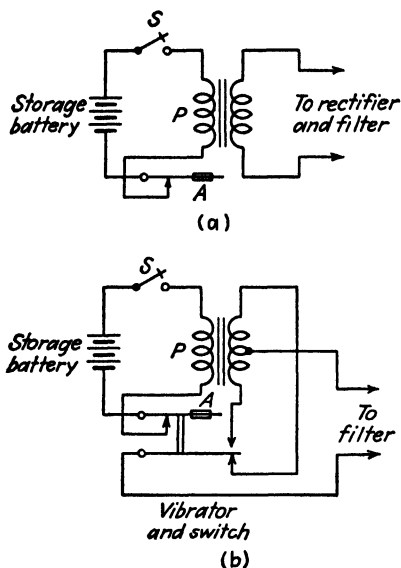



FIG. 154.—Circuits for providing high-voltage direct current from a low-voltage storage battery. In (*a*) the vibrator makes and breaks the primary circuit inducing a high voltage in the secondary of the transformer. This voltage is then rectified and filtered. When switch *S* is closed, current from the battery flows through primary *P*, attracting iron armature *A*, thus opening the circuit. The armature then returns to the original position and the cycle is repeated. In (*b*) the operation is similar, but as the vibrator makes and breaks the primary circuit, it also operates the mechanically connected reversing switch, which reverses the polarity so that no rectifier is needed.

Vibrators.—If only a radio receiver is to be powered, and if the requirements are low, then the direct voltage for the plates of the tubes would probably be obtained from some vibrator system. Here there are at least two possibilities: (*a*) a mechanical vibrator arranged as in Fig. 154*a* can be made to make and break the circuit in the primary, and with a suitable transformer, the desired high alternating voltage will be induced in the secondary. This voltage can be rectified by conventional power supplies that have been described in this chapter, and the rectified output is then available for the plates of the tubes. (*b*) A synchronous vibrator commutator can be used. As indicated in Fig. 154*b*, this makes and breaks the primary, and it also "switches" the secondary circuit in exact synchronism so that no rectifier tube is needed. In this connection it should be remembered that a vacuum-tube rectifier is in a sense a switching circuit that connects the load to the supply only when the polarity is such that the current flow always is in the same direction. Mechanical rectifiers of the nature shown in Fig. 154*b* have been used in electric circuits for many years.

SUMMARY

A rectifier is a device that converts alternating current into direct current. Current readily flows through a rectifier in one direction, but little or no current flows through it in the other direction. A rectifier unit is the entire rectifying apparatus including transformers, etc.

An ideal rectifier would pass current in one direction without opposition, but would pass no current in the other. Its impedance would be either zero or infinity, depending on the direction of current flow. When an alternating voltage is applied to a rectifier, current flows for the positive half cycle, but not for the negative half cycle. The current that flows and the rectified voltage consist of pulses. In a sense a rectifier is a distorter because it changes the wave shape.

The basic rectifier circuits used most widely are the half-wave and full-wave types. When working into a pure resistance load, or a filter circuit operating with choke input, the equation for the half-wave current or voltage is

$$e = \frac{E_m}{\pi} + \frac{E_m}{2} \sin 2\pi ft - \frac{2E_m}{3\pi} \cos 2\pi 2ft - \frac{2E_m}{15\pi} \cos 2\pi 4ft \cdots,$$

where the first term is the rectified direct current, and the other terms are alternating-current components, or harmonics. For the full-wave rectifier unit, the equation for the full-wave current or voltage is

$$e = \frac{2E_m}{\pi} - \frac{4E_m}{3\pi} \cos 2\pi 2ft - \frac{4E_m}{15\pi} \cos 2\pi 4ft - \frac{4E_m}{35\pi} \cos 2\pi 6ft \cdots,$$

where the terms have the significance just explained.

The rectifiers, that is, the units which have the one-way current-carrying characteristics, include crystals, devices with barrier layers, high-vacuum diodes, and gas diodes. The devices often have a high voltage across them on the negative, or nonconducting, part of the cycle. Their rated peak inverse voltage, which is the voltage they can stand in the negative direction, should not be exceeded.

The type of load used affects the rectifier output. With resistance only, or resistance and inductance only, the wave shapes of the currents and voltages are essentially unaffected by the nature of the load. If a capacitor is between the rectifier and load, the capacitor may cause the wave shape to be different from the simple half wave or full wave, in which event the equations previously given do not apply.

Usually the wave which results from rectification must be filtered so that the harmonics do not reach the load. The filters commonly used are the choke-input and condenser-input types. A filter connected with choke input may actually operate as condenser input if the choke is too small. To ensure choke-input operation, the input choke and load resistance have the approximate relation

$$L_1 = \frac{R}{1000} \quad \text{and} \quad R = 1000L_1.$$

If L_1 is less than this value, then cutout occurs during the rectification cycle, and the rectifier does not pass current. Then the condenser discharges through

the load, and the wave shape does not follow the equations previously given. If cutout occurs, the voltage rises on light loads and no loads, and the regulation is poor. Gas tubes should be used with choke-input operation, or the peak current may be excessive and tube damage may occur.

The inductance of iron-core choke coils varies with the amount of direct current being carried, and it is called the "incremental inductance." With such coils the inductance is high when the direct current is low, and vice versa. This is fortunate, because to prevent cutout less inductance is required when the load is drawing a large current.

Bleeders always should be used with rectifiers that have condensers with little leakage. A bleeder is a high resistance connected across the filter output. A bleeder discharges the condensers when the power supply is shut off, and thus is an important safety device. A bleeder also improves regulation, and can be made to be a voltage divider, making available several different output voltages. By-pass condensers should be connected across the sections of a voltage divider to prevent one connected load circuit from passing alternating-current signals into other connected loads.

Resistance-capacitance filters instead of inductance-capacitance filters are used often, particularly where plenty of rectified voltage is available and where the load current is reasonably low. It should be remembered that a series inductor will offer but little opposition to the flow of direct current, but a series resistor will offer much opposition and will have a large voltage drop across it.

Voltage stabilizers that use either high-vacuum thermionic tubes or cold-cathode tubes are used for stabilizing the voltage output of a rectifier. Ballast tubes also are used for this purpose. Voltage doublers put out a direct voltage that is about twice the peak value of the alternating voltage that is impressed across the rectifiers.

Polyphase rectifiers of the three-phase and six-phase types are used where large amounts of direct-current power must be delivered to some device, such as a large radio transmitter. Bridge rectifiers often are used.

The plate voltage for vacuum tubes sometimes is obtained by having storage batteries supply the primary of a transformer and the contacts of a mechanical vibrator in series. The vibrator contacts make and break the storage-battery circuit, resulting in an interrupted current and an induced voltage in the secondary. This may be rectified by tubes or by contacts driven by the vibrator.

REVIEW QUESTIONS

1. Explain how the definitions for rectifier and rectifier unit differ from the common usage of the terms in radio.

2. What would be the resistance characteristics of an ideal rectifier? Would there be a power loss in an ideal rectifier? Explain.

3. Why may rectification be considered a form of distortion? Should distortion always be avoided in radio circuits?

4. Why is a transformer often used with a rectifier?

5. What are the important components existing in the output of a half-wave rectifier? A full-wave rectifier? How may this be proved?

6. What is meant by a harmonic? Do these exist in rectifier outputs?

7. Name several devices that are used as rectifiers.

8. What is meant by inverse voltage?

9. What is the magnitude of the inverse voltage in half-wave and full-wave rectifiers. Why is it of importance?

10. Explain why the inverse voltage determines the maximum value of the direct voltage that a rectifier may deliver.

11. What is meant by cutout? What influence does it have on the wave shape of a rectifier output?

12. Explain how an inductor (choke coil) in series with the load produces "smoothing action."

13. In rectifier filters why are condensers never placed in series with the load?

14. If a condenser is in parallel with a resistance load, will it always produce smoothing action? Explain.

15. What units comprise a complete system for changing alternating to direct power?

16. If a direct-current milliammeter is connected in the output circuit of a rectifier, what value will it indicate?

17. What is meant by the terms "choke-input filter" and "condenser-input filter." Do these influence the rectifier output wave shape?

18. In selecting coils for a choke-input filter what precautions should be taken? What is meant by incremental inductance? How does it enter into filter design?

19. Why are gas tubes seldom used with condenser-input filters? What precautions should be taken if they are used?

20. What is a bleeder and why is it used? What is a voltage divider, and why is it used?

21. Explain what is meant by a resistance-capacitance filter? Discuss the conditions under which such filters are used.

22. What is a voltage stabilizer, and what types of tubes are used in them? What is a ballast lamp, and where is it used?

23. What is a voltage doubler, and what is its basic principle of operation?

24. What is a polyphase rectifier and where are they used in radio? What is an important advantage of such rectifiers?

25. Discuss several ways of operating remote radio equipment from storage batteries.

PROBLEMS

1. It is desired to construct a half-wave gas-tube rectifier that will give a direct voltage output of 300 volts. If the source voltage is 115-volts 60-cycles, what should be the approximate transformer turns ratio? Repeat for a full-wave rectifier.

2. Suppose that a full-wave rectifier is connected to an inductive load with negligible resistance. If the various voltage harmonics have the magnitudes indicated in Fig. 129, what will be the relative magnitudes of the harmonics in the current that flows. Let the lowest frequency be unity.

3. Repeat the calculations for the problem, starting on page 230, but with two 12-henry 200-ohm chokes in series.

4. A load varies from infinity to 2000 ohms resistance. If a 15-henry choke is to be used, what must be the resistance and power-handling capacity of a bleeder to prevent cutout?

5. Following the illustrative problem starting on page 240, design a power supply that will deliver 500 volts at 200 milliamperes.

6. Following the illustrative problem starting on page 243, design a power supply that will deliver 350 volts at 100 milliamperes.

7. Following the illustrative problem starting on page 246, redesign the voltage divider to provide a 270-volt 50 milliampere tap instead of one at 180 volts.

8. The resistance-capacitance filter of Fig. 144 is connected across a full-wave rectifier on which is impressed a maximum 60-cycle voltage of 500 volts. If R_1 is 5000 ohms and C_1 is a 16-microfarad electrolytic condenser, what will be the percentage ripple at a load resistance of $R = 7500$ ohms? Do you believe this would be sufficiently low if R were a small audio amplifier? What must be the power-handling capacity of R_1?

9. If the gas tube of Fig. 145 has a current through it of 15 milliamperes and a voltage across it of 150 volts, what will be its equivalent resistance? Will this remain constant for all current values? How much power is dissipated in the tube when operated at the values indicated above? If R equals 2500 ohms, how much power will be delivered to it?

10. Use the curve of Fig. 148, and calculate the resistance and power consumption of the ballast lamp at various points over the operating range. Plot these data as curves with impressed voltage on the X axis.

CHAPTER VIII

VOLTAGE AMPLIFIERS

Vacuum-tube amplifiers were discussed briefly in Chap. VI, where it was explained that there are two basic types, voltage amplifiers and power amplifiers. It is important that this fact be kept in mind.

Voltage amplifiers, as the name indicates, are for increasing or amplifying alternating signal voltages. In a typical case, the voltage impressed on the input of an amplifier may be of the order of 0.0001 volt, and this signal voltage may be increased to perhaps 50 volts so that it is sufficiently large to drive a power tube (page 196).

Power amplifiers, as the name indicates, are for amplifying power. This is not strictly correct, however, because many power amplifiers draw negligible power from the source of voltage driving them. Nevertheless the name is in general use to designate a device designed primarily to deliver signal power, for example, to a loudspeaker.

Of course voltage amplifiers put out some power, and power amplifiers often increase the signal voltage. In general, however, a voltage amplifier is designed for maximum signal amplification or gain, but a power amplifier is designed to deliver maximum undistorted power output. The objectives are different, and the design principles are different.

Voltage amplifiers will be considered in this chapter, and power amplifiers in the chapter that follows.

Signal Distortion.—As has been explained, the purpose of a voltage amplifier is to increase the alternating signal voltage that is impressed across its input. It usually is desired that the output signal voltage be an exact replica of the input signal voltage. This means that the **distortion,** defined as a change in wave form, must be negligible. Distortion is of three types, as will now be explained.

Frequency distortion[1] is that form of distortion in which the change is in the relative magnitudes of the different frequency components of a wave, provided that the change is not caused by nonlinear distortion. Thus suppose that a complex signal wave is impressed on the input of an amplifier. If the amplifier is so designed that it increases some of the frequency components of the complex signal wave more than it increases others, then frequency distortion results. Or if a complex signal wave containing many components is impressed on a transmission circuit, such as a line or a network, and if the circuit offers more attenuation to certain frequencies than to others, then the components of various frequencies will not arrive at the receiving end with the same relative magnitudes that they had at the sending end, and frequency distortion results. This sometimes is called "amplitude distortion."

When a "frequency run" is made on a piece of equipment to determine the "frequency response," the frequency distortion is being measured. Thus if the response characteristic curve of an audio-frequency voltage amplifier is essentially "flat" from 50 to 10,000 cycles, the frequency distortion is negligible over that range.

Nonlinear distortion is the form of distortion that occurs when the ratio of voltage to current (the impedance), using effective values, is a function of the magnitude of either the voltage or the current. As an example, suppose that an amplifier is designed for an input voltage that should not exceed 0.01 volt. If a sine-wave voltage of 0.01 volt or less is impressed on the amplifier, the output voltage will be approximately a sine wave. On the other hand, if a sine-wave voltage greater than 0.01 volt is impressed on the amplifier, it will be overloaded, and the output will be not a sine wave, but a complex voltage wave containing sine-wave components of various frequencies. Or suppose that an iron-cored transformer is designed to work at a certain volume level, that is, strength of signal. If a sine-wave voltage below the maximum rating is impressed on this transformer, the output will be approximately a sine wave, but if the impressed voltage is too great, then the transformer will be overloaded, the magnetic flux density in the core will be excessive, and distortion will result.

[1] The definitions given here essentially are as listed in American Standard Definitions of Electrical Terms, American Institute of Electrical Engineers, 1941.

It is possible to obtain a qualitative idea of the distortion caused by an amplifier, by a transformer, or by other devices, by observing the input and output signal waves on a cathode-ray oscilloscope (page 565). It is a simple matter to determine quantitatively the amount of distortion if a **wave analyzer** is available. This device "tunes through" the frequency spectrum covered by the complex output wave, and gives the frequency and magnitude of each alternating-current component that was created by the distortion process. The amount of distortion that can be tolerated depends on circumstances. In audio-frequency circuits it is common practice to hold the nonlinear distortion to about 5 per cent or less, because experience shows that the ear cannot detect less than about 5 per cent distortion. Nonlinear distortion is quite likely to occur in power amplifiers because usually these are designed to operate the tubes at their upper power-handling limit.

Delay distortion is that form of distortion in which the time of transmission, or the delay, encountered by a signal in passing through a network varies with frequency. As an example of delay distortion, consider the audio-frequency transmission line. The velocity of propagation given on page 135 was for a 1000-cycle signal, and it is different at other frequencies. When a complex speech wave containing many frequency components, each having a *different phase relation* with respect to the other components, is impressed on the transmission line, some components will be delayed more than others. At the receiving end the various frequency components will not have the same phase relations that they had at the sending end. Thus delay distortion has resulted. A similar phenomenon exists in an amplifier. If the phase relation between the *input voltage and output voltage* is measured at various frequencies, it will be found that the phase relations between the voltage components is not the same. A low-frequency component may be shifted through a certain angle; yet an intermediate-frequency component may be shifted a negligible amount. Note again that this shift is between the same quantity (say voltage) measured at two different points. Such an amplifier would have delay distortion, or phase distortion, as it often is called.

In circuits used for speech only, delay distortion does not result in serious impairment of speech quality or intelligibility. The ear normally receives sounds by various paths, some of which are longer than others, and as a consequence phase shifts exist. In

the transmission of photographs by electrical means, and in television, delay distortion and the resulting phase shift is of vital importance. If delay distortion exists, then the received image will not be a replica of the image transmitted.

Classification of Voltage Amplifiers.—There are several ways in which voltage amplifiers may be classified, one basis being the frequencies that are to be amplified. Thus there are **audio-frequency amplifiers** that amplify frequencies from about 50 to 10,000 cycles (page 41), and there are **radio-frequency amplifiers** that amplify frequencies above this range. Also, there are **video amplifiers** used in television that amplify frequencies from about 30 to 3,000,000 cycles or more, depending on the design.

Another method of classification is based on the manner in which the tubes of an amplifier are connected together for passing the signal from one tube to another. More than one tube often is required to give the necessary amplification, and thus two or more tubes or stages of amplification are interconnected, each giving some amplification and feeding the increased signal into the next. The interstage coupling circuits used are of several types, the most common being a resistance-capacitance network or an interstage transformer. The classifications are, therefore, the **resistance-capacitance coupled amplifier,** usually shortened to **resistance-coupled amplifier,** and the **transformer-coupled amplifier.** Audio-frequency amplifiers usually are resistance-coupled, although coupling through interstage audio-frequency transformers is used. Radio-frequency voltage-amplifying stages usually are coupled through tuned radio-frequency transformers. Tuning is used because in radio reception usually it is desired to pass a certain signal band and reject all others.

The Vacuum Tube as a Voltage Amplifier.—This is an important application of the vacuum tube. Other devices, such as barrier-layer rectifiers and crystals, offer the vacuum tube some competition for such purposes as rectification (including modulation and demodulation), but for amplification, the vacuum tube has little competition. (It is interesting to note that a telephone receiver directly connected mechanically to a carbon-granule telephone transmitter has amplifying properties and was once employed to a limited extent as an amplifier in long telephone lines. This receiver-transmitter combination is used now in some hearing aids.)

Triodes, tetrodes, and pentodes all are used as voltage amplifiers.

Because of their high amplification factor and the resulting high amplification (or voltage gain) per stage, pentodes probably are the most widely used in *audio-frequency* voltage amplifiers. Tetrodes and particularly pentodes are most widely used in *radio-frequency* voltage amplifiers. There are several reasons for this: (*a*) Their high amplification is an advantage; (*b*) and possibly a more important reason, because of the shielding afforded by the suppressor grid and screen grid between the plate and control grid, there is negligible feedback of signal from plate to control grid and little tendency for self-oscillation except at the high radio frequencies. This will be considered in detail in the next chapter. The basic principles of the voltage amplifier will be discussed, using a triode as an amplifier because of the simplicity of this tube.

When a vacuum tube is used as an amplifier, *both* the tube and the load in the plate circuit determine the amount of amplification or voltage gain per stage. The characteristic curves of Chap. VI were for vacuum tubes *alone*, because that chapter concentrated on tubes as electronic devices. Such curves are called "static characteristic curves," and it is of importance to note that such curves were taken *without* resistance (or impedance) in the plate circuit. But when vacuum tubes are used for amplifiers, resistance (or impedance) must be connected in the plate circuit. Thus the static curves alone are not able to show the plate-current variations that occur when an alternating signal voltage is impressed on the grid. For this purpose *dynamic characteristic curves* must be used. These curves are the combined characteristics of the tube and its connected load. The methods of measuring and calculating these dynamic curves will be considered on page 324.

The basic circuit for a triode amplifier, and the dynamic characteristic curve for this triode with certain applied electrode voltages and plate-load resistance, is shown in Fig. 155. The source of grid-bias voltage is E_{cc}. Because the grid is negative, the grid current is assumed zero, and all this voltage appears between the grid and cathode as a grid-bias voltage E_c.[1] This fixes the point of operation on the dynamic characteristic curve. The plate-supply voltage E_{bb} causes plate-current flow I_{bo} when the tube is in the quiescent state with no grid signal voltage applied. This

[1] The letter symbols E_{cc}, E_c, etc., are in accordance with Standards on Electronics—Definitions of Terms, Symbols, The Institute of Radio Engineers, 1938.

current I_{bo} flowing through the load resistor R_L causes a direct voltage drop $E_L = I_{bo}R_L$, and the actual direct voltage on the plate is

$$E_b = E_{bb} - I_{bo}R_L. \tag{77}$$

Now when an alternating signal voltage having an instantaneous value of e_g and an effective value of E_g is impressed on the grid of the tube, the total instantaneous grid-voltage variations e_c are the sum of the direct grid bias E_c and the instantaneous alternating variations e_g. (Here a pure sine wave is used for convenience to represent the complex signal voltage that in practice would be impressed for amplification on the grid of the tube.) As a result, the grid voltage is driven less negative and more negative, and the total instantaneous plate current is as shown by i_b. This plate current contains several components, those of present interest being an average, or direct-current, value I_b and an alternating current of instantaneous value i_p. If the dynamic curve *is a straight line*, then I_b, the average value of direct plate current that flows with applied grid signal voltage, will be the same as I_{bo} the quiescent value of plate current that flows without applied grid signal voltage. Also, if the dynamic curve is a straight line, the alternating-current component of plate current i_p will be a pure sine wave. But the dynamic curve is not a straight line (although it may approximate a straight line), and hence I_b is greater than I_{bo}, and i_p is not exactly a sine wave, but is a complex wave containing harmonics, or distortion components. The curvature of the dynamic curve of Fig. 155 has been accentuated to illustrate the difference between I_b and I_{bo}.

It has been shown that an alternating grid signal voltage causes a corresponding alternating plate-current flow. *But, the purpose of a voltage amplifier is to amplify or increase voltage.* For this reason the alternating plate-current component must be passed through some impedance so that the plate-current variations will cause a corresponding alternating voltage. Then, the amplifier, when driven by weak alternating signal voltage, will put out an increased alternating signal voltage, giving an over-all voltage gain.

Equivalent Circuit of a Vacuum-tube Amplifier.—In the preceding section it was explained that an alternating-current component flowed in the plate circuit of a vacuum tube, and through the plate load resistor, when a signal voltage was applied between the grid and cathode of a vacuum tube. It is true that the effect of the

grid signal voltage is felt in the plate circuit, yet in a sense, these two circuits are independent. Applying the ordinary concept of current flow to the plate circuit, alternating signal current flows

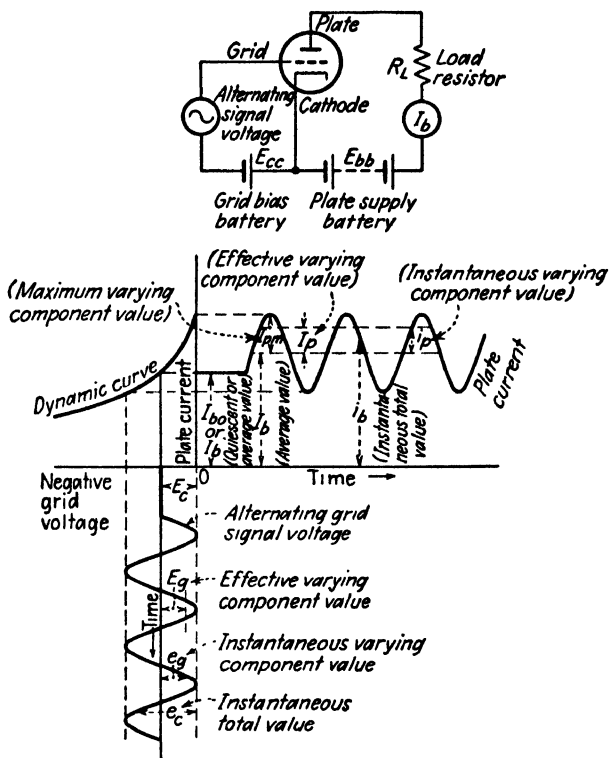



FIG. 155.—The triode as an amplifier with a load resistor in the plate circuit operates as shown. The dynamic curve usually is much straighter than indicated. The curvature has been accentuated to indicate more clearly the difference between I_b and I_{b_0}. The major letter symbols to be used with vacuum tubes are indicated in this figure. These are in accordance with the Standards of the Institute of Radio Engineers. Briefly, c or g applies to the grid, and b or P applies to the plate. Also, i or e applies to instantaneous values of current or voltage, and I or E applies to quiescent, average, or effective values.

in the plate circuit because an alternating signal voltage forces this current through the circuit.

The question arises: Where is the alternating voltage that forces the alternating current through the plate circuit? The answer is, of course, that there is no such alternating voltage, the grid merely permits more or less plate current to flow as the grid voltage varies in accordance with the applied alternating grid signal voltage. But

for convenience in amplifier design, it is helpful to represent the vacuum tube as an ordinary circuit in which a *fictitious or imaginary voltage* applied in series with the alternating plate resistance of the tube (page 186) and the external load resistance forces the same alternating signal current through the circuit as flows when the grid signal voltage is applied. This gives the circuit of Fig. 156 as the electric circuit that is equivalent to a vacuum tube acting as an amplifier.

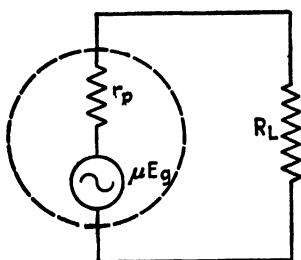



Fig. 156.—The vacuum tube as a linear or nondistorting amplifier acts as a generator of open-circuit voltage μE_g and internal resistance r_p, where μ is the amplification factor of the tube, r_p is the plate resistance of the tube, and E_g is the alternating signal voltage impressed between grid and cathode. This signal voltage often is represented by E_s.

It will be noted that the fictitious or imaginary voltage in this equivalent circuit has the magnitude μE_g, where μ is the amplification factor of the tube and E_g is the effective value of the applied alternating grid signal voltage. The reason that μ appears is because the grid is μ times as effective as the plate in controlling the electron current flow. Thus if a voltage of E_g volts is applied in the grid circuit, a voltage of μE_g volts would have to be applied in series between the *plate and cathode* to produce the same plate current flow that the grid voltage E_g causes. Of course the alternating plate current must flow through the alternating plate-circuit resistance r_p, because this is in series with the load resistance R_L.

It is now possible to write the equations for the simple vacuum-tube amplifier of Fig. 155. Based on the equivalent amplifier circuit of Fig. 156, and following the theory of a series circuit

$$I_p = \frac{\mu E_g}{r_p + R_L}, \qquad (78)$$

where I_p is the effective value of the alternating signal current in amperes that flows in the plate circuit and through the load resistor R_L when μ is the amplification factor of the tube, E_g is the effective value of the grid signal voltage in volts, and r_p and R_L are the internal plate resistance and external load resistance in ohms.

The magnitude of the voltage across the load resistance is of vital importance, because this is the amplified signal output volt-

age. This voltage may be utilized for some purpose or may be impressed on a following stage for further amplification. This amplified output signal voltage E_L existing across R_L will of course be the I_pR_L voltage drop, or

$$E_L = I_pR_L = \frac{\mu E_g R_L}{r_p + R_L}, \tag{79}$$

and will be in volts when I_p is in amperes and R_L is in ohms. There will be an I_pr_p drop *within* the tube, but this is wasted because it is inside the tube and cannot be utilized for external purposes.

As has been repeatedly stressed, *voltage* amplifiers are being considered in this chapter. Of particular importance, therefore, is the voltage gain or amplification per stage. This will be the ratio of the output voltage to the input voltage. From Eq. (79)

$$A_v = \frac{E_L}{E_g} = \frac{\mu R_L}{r_p + R_L}, \tag{80}$$

where A_v is the voltage gain.

Illustrative Problem.—The triode portion of a type 6SQ7 vacuum tube has an amplification factor of 100 and a plate resistance of 91,000 ohms, and is operated with a plate load resistor of 250,000 ohms. The effective value of the applied grid signal voltage is 0.1 volt. Calculate the effective value of the plate current, the output voltage drop across R_L, and the voltage gain for the stage, expressed as a ratio, and in decibels.

Solution.—Step 1. From Eq. (78), the effective value of the plate current is $I_p = \mu E_g/(r_p + R_L) = 100 \times 0.1/(91{,}000 + 250{,}000) = 10/341{,}000 =$ 0.00002935 ampere.

Step 2. From Eq. (79), the voltage drop across load resistor R_L is $E_L = I_pR_L = 0.00002935 \times 250{,}000 = 7.33$ volts. Or, $E_L = \mu E_g R_L/(r_p + R_L) = 100 \times 0.1 \times 250{,}000/(91{,}000 + 250{,}000) = 7.33$ volts.

Step 3. From Eq. (80), the voltage gain for this tube and the associated circuit is $A_v = E_L/E_g = 7.33/0.1 = 73.3$ or $A_v = \mu R_L/(r_p + R_L) = 100 \times 250{,}000/(91{,}000 + 250{,}000) = 73.3$.

Step 4. From Eq. (28), page 96, the voltage gain in decibels is $n = 20 \log_{10} E_L/E_g = 20 \log_{10} 73.3 = 20 \times 1.865 = 37.3$ decibels.

If two of these stages are connected together so that one feeds into the other and the complete gain of each is realized, then the over-all voltage gain of the two stages is $73.3 \times 73.3 = 5370$. The gain in decibels for the two stages is $37.3 + 37.3 = 74.6$ decibels. To check this, $n = 20 \log_{10} A_v = 20 \log_{10} 5370 = 20 \times 3.73 = 74.6$ decibels. Thus in amplifiers of more than one stage the gains per

stage are *multiplied* together to give the over-all gain when the gains are expressed as ratios, and the gains per stage are *added* together to give the over-all gain when the gains are expressed in decibels.

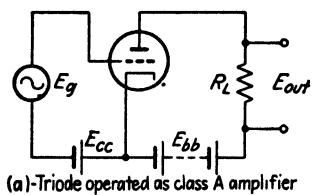
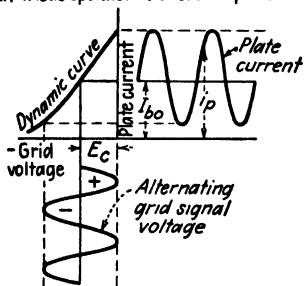
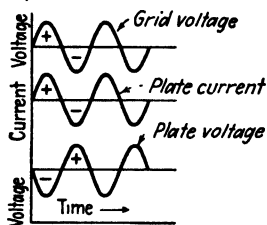



FIG. 157.—Diagrams showing the phase relations in an amplifier with a load of pure resistance. From diagram (*c*) it is evident that the grid voltage and plate current are in phase, but that the plate voltage is 180° out of phase with the grid voltage and plate current.

The simple amplifier circuit of Fig. 155 and the equivalent circuit of Fig. 156 may appear to be unrelated to radio. This is far from the truth. These circuits are of basic importance to any amplifier, and they illustrate the principles involved. Of course in radio, tuned parallel circuits are in the plate circuit, but it will be recalled that the input impedance of a tuned parallel circuit *is pure resistance at resonance.* Thus the figures here discussed and the equations derived are of basic importance in studying radio-frequency amplification.

Phase Relations in Amplifiers with Resistance Loads.—As was mentioned in discussing delay distortion, such matters as phase relations usually are of little consequence when speech only is being transmitted. It was explained, however, that in video amplifiers used in television it is important that no delay distortion occurs. There are other instances, such as in the design of feed-back amplifiers and oscillators, when the phase relations between the amplifier input and output voltages must be known. This section will consider such relations for amplifiers with resistance loads, which, it has been stressed, are the basic type. The discussion will not apply at the very high radio frequencies, above perhaps 100,000,000 cycles, depending on the type of the tube used (see page 290).

The basic amplifier circuit and the curves of Fig. 155 have been reproduced in Fig. 157. Also shown in this figure are the instan-

taneous phase relations of grid voltage, plate current, and plate voltage. These relations are explained as follows: When the positive half cycle of the alternating signal voltage between the cathode and grid occurs, the total grid voltage is made less negative than direct grid-bias value E_c, and this permits an increase in the plate-current flow, producing the positive half cycle of plate current. Thus the alternating plate current is in phase with the grid signal voltage. Now the instantaneous plate voltage, as measured between the cathode and plate, is equal to the applied voltage E_{bb} minus the instantaneous drop in voltage in the load resistor. That is

$$e_p = E_{bb} - i_p R_L. \tag{81}$$

Thus, when i_p *increases*, the voltage drop $i_p R_L$ across the load resistor increases, and e_p *decreases*. For the negative half cycle of

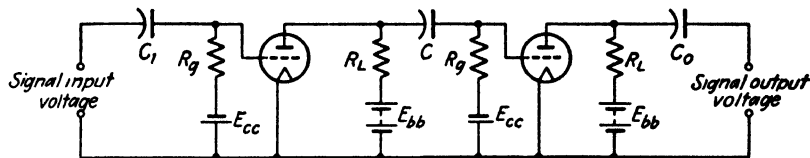



FIG. 158.—Simplified diagram of a resistance-coupled audio-frequency amplifier. In an actual amplifier the voltages indicated would probably be supplied by one power supply or by one set of batteries.

applied grid voltage the opposite action occurs. Thus the voltage e_p from cathode to plate is 180° out of phase with the plate current and the grid signal voltage. The alternating voltage from cathode to plate is the amplified output voltage. Attention is again called to the fact that this statement applies *only* for a load of pure resistance and at the frequencies specified.

Resistance-coupled Audio-frequency Amplifiers.—The output signal voltage that exists across the load resistor of a single-stage amplifier is given by Eq. (79). As has been mentioned, one stage may produce insufficient gain for a given application, and two or more stages, coupled together so that the amplified output voltage of one feeds into another for further amplification, may be necessary. One widely used method of interconnection is shown in Fig. 158 and will be considered now.

If the grid of the second tube were connected directly across R_L, then the grid would have a positive voltage on it. For amplification, the grid should be biased negatively so that the tube will operate as indicated in Figs. 155 and 157. Thus the **coupling**

capacitor C is inserted to isolate the grid of the second tube from the positive plate voltage of the first tube. The **grid resistor** R_g is used so that the biasing voltage E_{cc} of the second tube will be impressed on the grid. If the grid of the second tube is left "free" (that is, if coupling capacitor C is inserted but R_g is made infinite, or open), then the grid of the second tube will pick up negative electrons, will acquire a negative potential that may be sufficient to prevent electron current flow to the plate, and thus will "block" the tube. Grid resistor R_g sometimes is referred to as a grid leak, and of course it does prevent the grid from accumulating a charge, but actually it does it by impressing the negative voltage of the

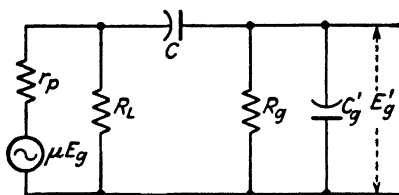



FIG. 159.—Circuit equivalent of the first tube and coupling circuit of Fig. 158.

battery E_{cc} on the grid so that the grid is at the correct fixed negative potential. Then, the tube operates on the straight part of its characteristic curve.

A single stage of amplification, such as that shown in Figs. 155 and 157, theoretically will operate at any frequency from zero to infinity. Actually, the upper limit would be determined by the stray wiring capacitance, residual capacitance of R_L, and the high-frequency characteristics of the tube. For the circuit of Fig. 158, the situation is different; this circuit has definite low-frequency and high-frequency limitations. The *low-frequency* limitations are the result of capacitor C, but the *high-frequency* limitations are caused by the input capacitance of the grid circuit of the second tube and by the stray wiring capacitance.

These effects can be explained by the use of Fig. 159, which is an equivalent circuit for the first tube and the coupling network. It is this part of Fig. 158 that largely determines the frequency characteristics of the amplifier. (Of course the condenser in the output circuit of the second tube of Fig. 158 would have an effect if *current* were drawn, but if voltage only were supplied by the amplifier the effect of this condenser would be negligible.) In this circuit C'_g represents the input capacitance of the second tube and the wiring

capacitances of the grid circuit. The effects of the grid-input capacitance of the first tube are neglected because it is assumed that the source of alternating signal voltage E_g holds this voltage at any value desired. In other words, in this analysis attention is concentrated on the *interstage coupling circuit.*

The problem is this: to be able to determine at any frequency the output voltage that is impressed by one stage on the grid of the second tube. Now Fig. 159 is an ordinary series-parallel circuit. The output voltage will be the voltage drop across condenser C_g', which will be the current through C_g' times the reactance of C_g'. Thus it is possible to write one equation in terms of the various values of frequency, r_p, R_L, etc., and by substituting in this equation, the magnitude of the output voltage can be found. Also, such an equation will give the phase angle between the input and output voltages. This method is not too difficult, and in fact, everything considered, it may be the simplest means; but it is not used in practice. Instead, the equivalent circuit of Fig. 159 is considered at three frequencies, and is further simplified into three equivalent circuits, one applying at intermediate frequencies, one at low frequencies, and one at high frequencies.

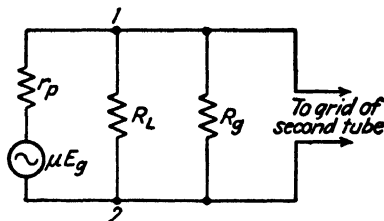



Fig. 160.—Equivalent circuit of the first tube and coupling circuit of Fig. 158 at *intermediate* audio frequencies.

Equivalent Circuit at Intermediate Frequencies.—If the capacitive reactance of a *series* capacitor, such as C of Fig. 159, is *low* in comparison with the other elements of the circuit, then the effect of such a series capacitor can be neglected. If the capacitive reactance of a *parallel* capacitor, such as C_g' of Fig. 159, is *high* in comparison with the other elements of the circuit, then the effect of such a parallel capacitor can be neglected. Although the numerical values will not be presented here, but will be given in the following section, this situation exists at intermediate frequencies, such as 1000 cycles, in a typical resistance-coupled audio-frequency amplifier. Thus at intermediate frequencies the circuit of Fig. 159 can be simplified into that of Fig. 160, in which C and C_g' have been omitted because at 1000 cycles their effects are negligible. This circuit then follows the theory of Fig. 156 and the same equations apply with slight modification. Thus, R_L of Fig. 156 and

Eqs. (79) and (80) is the equivalent parallel resistance of R_L and R_g of Fig. 160. This point will be clarified on page 286. The phase shift for the circuit of Fig. 160 will be 180°, just as for Fig. 157.

Equivalent Circuit at Low Frequencies.—At low frequencies, say below about 50 cycles, the reactances of the two condensers of Fig. 159 are quite high. There is, accordingly, even more justification for neglecting the *parallel* capacitor C_g' than before. But the effect of a high *series* reactance cannot be neglected, and, hence, C appears in the equivalent circuit of Fig. 161 that represents the circuit of Fig. 159 at low frequencies. The output voltage of this circuit is the IR drop across R_g. The current that flows to R_g is affected considerably by the capacitive reactance of condenser C. At low frequencies the capacitive reactance will be very high, and the current flow to R_g and the voltage drop across R_g will be low. Hence, for given values of r_p, R_L, and R_g, the voltage output at low frequencies and the frequency distortion are determined by the capacitance of condenser C. As the frequency approaches zero, the reactance of condenser C approaches infinity, and thus negligible signal gets to the right of condenser C. It is possible to write one equation that will consider such factors, the equation being

FIG. 161.—Equivalent circuit of the first tube and coupling circuit of Fig. 158 at *low* audio frequencies.

$$\text{Low-frequency characteristics} = \frac{\mu R_L R_g}{(r_p R_g + R_L R_g + r_p R_L) - j\left(\frac{R_L + r_p}{2\pi f C}\right)}$$
$$= A_v \underline{/+\theta}. \tag{82}$$

All resistances are in ohms, and C is in farads. When this equation is used, A_v will be the magnitude of the voltage gain at frequency f cycles per second, and $+\theta$ will be the phase shift in degree *caused by the coupling* circuit. This must be added to the 180° shift, caused when the tube works into a load of pure resistance. This gives the over-all phase shift of the tube and coupling network. Sometimes a minus sign is written in front of μE_g to indicate this 180° phase shift. For advanced work this may be advisable, but it is confusing. For the purposes at hand, it is better to remember

that there is a basic 180° shift and that when θ is positive the two are added to obtain the over-all shift.

The derivation of Eq. (82) is not difficult and is based on simple series-parallel theory (page 84). The steps are as follows: First find the series impedance of C and R_g. Then find the equivalent series impedance of the parallel circuit between points 1–2 of Fig. 161. Combine this with r_p to find the total equivalent impedance into which the voltage μE_g works. Find the current through r_p by dividing μE_g by the total equivalent impedance. Next find the voltage drop across r_p and subtract this from the total applied voltage μE_g. This gives the voltage across points 1–2. The voltage across points 1–2 divided by the impedance of C and R_g in series will give the current through R_g. This current multiplied by R_g gives the voltage output, and the voltage output divided by the voltage input gives Eq. (82). Of course the vector nature of all values must be considered.

In amplifier design the response characteristics (that is, voltage gain plotted against frequency on the X axis) is not made "flat" over the entire band to be passed. Thus if an amplifier is to pass from 50 to 10,000 cycles, it usually is considered satisfactory if the gain at 50 cycles and 10,000 cycles is about 70 per cent of that at an intermediate frequency of say 1000 cycles. This reduction represents a loss of about 3 decibels. The corresponding phase shift for the coupling network is $\theta = 45°$. The frequency at which this occurs can be found from the denominator of Eq. (82) by equating the in-phase resistance term equal to the out-of-phase reactance term, so that $\theta = 45°$, and solving for f. The equation is

$$\text{Low frequency at which gain is down to 70 per cent of the intermediate value and phase shift is } 45° = \frac{R_L + r_p}{2\pi C\,(r_p R_g + R_L R_g + r_p R_L)}, \qquad (83)$$

where the resistances are in ohms and C is in farads. The reactance term of Eq. (82) is the part preceded by the letter j.

Equivalent Circuit at High Frequencies.—At high frequencies the capacitive reactances of condensers C and C'_g become low. Thus at high frequencies the effect of the series capacitor C can be neglected but the effect of the parallel capacitor C'_g must be considered. This gives the high-frequency equivalent circuit of Fig.

162. The output voltage is the *IX* drop across condenser C'_g. At high frequencies where the reactance of this condenser approaches zero, the voltage drop and, hence, the output voltage approach zero. Thus for given values of r_p, R_L, etc., the capacitance of condenser C'_g controls the high-frequency characteristics of the amplifier. An equation which applies at high frequencies and which is derived by the same general method as used for Eq. (82) is

$$\text{High-frequency characteristics} = \frac{\mu R_L R_g}{(r_p R_g + R_L R_g + r_p R_L) + j\,(2\pi f C'_g\, R_L R_g r_p)} = A_v \angle{-\theta}. \tag{84}$$

All resistances are in ohms, and C is in farads. When this equation is used, A_v will be the magnitude of the voltage gain at frequency

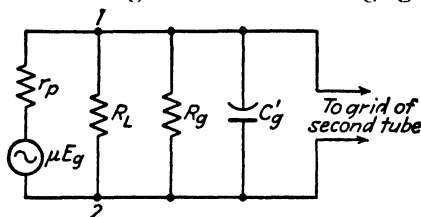



FIG. 162.—Equivalent circuit of the first tube and coupling circuit of Fig. 158 at *high* audio frequencies.

f cycles per second, and θ will be the phase shift *caused by the coupling circuit.* This must be subtracted from the 180° caused when the tube works into a load of pure resistance. This gives the over-all phase shift of the tube and coupling network.

As for the low-frequency case, no attempt is made in most instances to obtain a flat response curve over the frequency band to be amplified. The upper limit of usefulness for most audio-frequency purposes is considered to be when the gain is about 70 per cent of the value at intermediate frequencies, and the phase shift is 45°. The frequency at which this occurs is found from the denominator of Eq. (84) by equating the in-phase resistance term equal to the out-of-phase reactance term, so that $\theta = 45°$, and solving for f. The equation is

$$\begin{array}{l}\text{High frequency at which gain is}\\ \quad\text{down to 70 per cent of the}\\ \quad\text{intermediate value and phase}\\ \quad\text{shift is 45°}\end{array} = \frac{r_p R_g + R_L R_g + r_p R_L}{2\pi C'_g\, R_L R_g\, r_p}, \tag{85}$$

where the resistances are in ohms and C_g' is in farads. The reactance term of Eq. (84) is the part preceded by the letter j.

The basic equations for the design of resistance-coupled amplifiers have been given. The next several pages will explain how these equations are used in determining the sizes of the resistors and capacitors that should be used for typical amplifiers. Because the methods employed depend upon the types of tubes used, two examples will be considered; one will be an amplifier using triodes, and the other will be an amplifier using pentodes. Before these specific examples are considered, it is necessary to discuss an important topic: the method of obtaining maximum voltage output.

Maximum Output Voltage Considerations.—The purpose of a voltage amplifier is to take a feeble signal voltage from some device, such as a microphone, and amplify this feeble voltage until it is sufficiently large to accomplish something, for instance, until it is of sufficient magnitude to drive the grid of a power output tube that furnishes power to a loudspeaker. Thus in the design of a voltage amplifier, it is of importance to select resistors, etc., so that the maximum voltage gain per stage will be realized in so far as it is practicable to do so.

The equivalent circuit for a single stage of amplification is shown in Fig. 156, and Eq. (80) is the basic equation for the voltage gain of a single stage. For a given tube, and at specified conditions of operation (that is, certain electrode potentials), the values of μ and r_p are fixed. The matter to be determined is this: For specified conditions of operation, what value of load resistor R_L will give maximum amplified voltage output and maximum voltage gain per stage?

To determine this relation, different values of R_L/r_p will be taken, and will be substituted in Eq. (80). First, it will be assumed that $R_L = r_p$. Making these substitutions gives

$$A_v = \frac{\mu R_L}{r_p + R_L} = \frac{\mu r_p}{r_p + r_p} = \frac{\mu r_p}{2r_p} = \frac{\mu}{2} = 0.5\mu.$$

Thus when $R_L = r_p$, and these values are substituted in Eq. (80), it is found that the theoretical voltage gain $A_v = 0.5\mu$; that is, if the μ of the tube is 100, the gain per stage is $0.5 \times 100 = 50$, and if the μ of the tube is 13.8, the gain is 6.9. This is reasonable because an examination of Fig. 156 will show that if $R_L = r_p$, then

one-half of the available amplified voltage μE_g will be lost as an $I_p r_p$ drop *within* the tube, and the other half $I_p R_L$ will be delivered to the output terminals for useful purposes.

To continue the study of the relation between R_L and r_p, let $R_L/r_p = 2$, or $R_L = 2r_p$. Making the substitutions in Eq. (80) gives $A_v = \mu 2r_p/(r_p + 2r_p) = \mu 2r_p/3r_p = 2\mu/3 = 0.67\mu$. Thus when $R_L = 2r_p$, the theoretical gain is 67 per cent of μ. Similar calculations show that when $R_L/r_p = 3$, $A_g = 0.75\mu$; that when $R_L/r_p = 4$, $A_v = 0.8\mu$; etc. When these data are plotted, Fig. 163 results. This curve is very interesting because it shows (*a*) that the maximum possible voltage gain per stage equals the amplification factor μ of the tube, and (*b*) that to obtain this maximum possible gain, the load resistor R_L should equal infinity. Again, this is what would be expected from an examination of Fig. 156. If $R_L = \infty$, then no current I_p would flow. Then, there would be no $I_p r_p$ drop within the tube, and all the amplified voltage μE_g would appear at the output terminals.

Now it is very convenient to know these facts, but there are certain *very* practical considerations that must be examined. There is an average or direct-current value I_b of Fig. 155 that also must flow through the load resistor R_L. Obviously, R_L cannot equal infinity, because if it did the plate circuit would be open and no direct plate current would flow. The next thing to consider is what will happen if the ratio of R_L/r_p is some finite value such as, say 100. Then, for a tube with a plate resistance of $r_p = 12{,}000$ ohms, $R_L = 1{,}200{,}000$ ohms. If only 2.5 milliamperes direct current flowed in such a circuit, the direct-voltage drop across the load resistor would be $E_L = I_b R_L = 0.0025 \times 1{,}200{,}000 = 3000$ volts, and the power-supply voltage would have to be greater than this. Of course it would be impractical to use such voltage in an ordinary amplifier.

Suppose a value of $R_L/r_p = 5$ is considered. Then if $r_p = 12{,}000$ ohms, $R_L = 60{,}000$ ohms, and if $I_b = 2.5$ milliamperes, $E_L = 0.0025 \times 60{,}000 = 150$ volts. This is entirely reasonable, because the power supply would need to provide only 150 volts drop for the load resistor and 100 volts for the plate of the tube, totaling 250 volts [see Eq. (77) and accompanying discussion].

With $R_L/r_p = 5$, then a calculation, or Fig. 163, shows that 1/6 of μE_g would be lost within the tube and 5/6 of μE_g, or 83 per cent, would appear across the load resistor for use in the external circuit.

Now this is *very important.* Note that the shape of the curve of Fig. 163 is such that little additional voltage gain is obtained in going beyond $R_L/r_p = 5$; yet, if higher load resistances are used, the direct voltage drop becomes prohibitively large.

For these reasons, *with triodes,* values of the load resistance R_L are usually from three to five times the plate resistance r_p. It becomes more practicable to use added stages if more gain is necessary, rather than to try to utilize the full μE_g of the tube by using very large values of R_L.

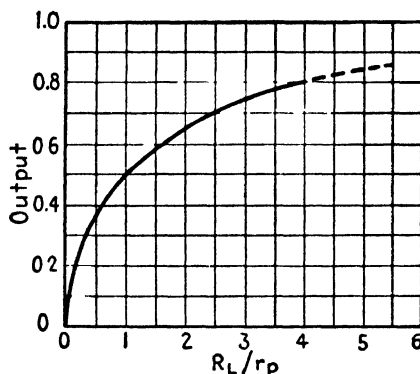



FIG. 163.—Curve showing the relative amplification of a single-stage resistance-coupled amplifier for various relations of R_L/r_p as compared to the maximum possible amplification, which is the amplification factor μ of the tube.

Attention is called to the fact that Fig. 163 and the accompanying reasoning apply to triodes, tetrodes, or pentodes *in theory.* The numerical values given in the preceding paragraph were *for triodes. They do not apply to tetrodes and pentodes.* For these tubes, the plate resistances are so high that ratios of $R_L/r_p = 1.0$, or 0.5, or 0.25 are more common. This will be discussed on page 291.

Resistance-coupled Voltage Amplifiers Using Triodes.—There are two aspects to the design of a resistance-coupled amplifier: (*a*) The circuit arrangement and the values of the resistors, etc., must be such that the proper direct voltages are between the tube electrodes and the cathode; (*b*) the circuit and the values of resistors and capacitors must be such that the applied alternating signal passes through the amplifier and that the desired voltage amplification results. Of course, experience is of much value in the design of an amplifier, and in the following example existing practices will be used.

Illustrative Problem.—It is desired to design an audio-frequency resistance-coupled voltage amplifier which will have a variable gain up to about 40 decibels, which will be substantially flat from 50 to 10,000 cycles, and which uses triodes.

Solution.—Step 1. General Aspects. A voltage gain of 40 decibels is an amplification ratio of 100 (Fig. 56, page 97). An examination of a tube manual shows that the maximum amplification factor of a standard triode is 100. Since all the theoretical gain cannot be utilized, two triodes must be used. If two are used, then each stage must have a gain of 20 decibels, or a voltage ratio of 10. If a ratio of $R_L/r_p = 5$ is used, then 5/6, or 83 per cent, of the theoretical amplification will be utilized. This means that the amplification factor must be at least $10/0.83 = 12$.

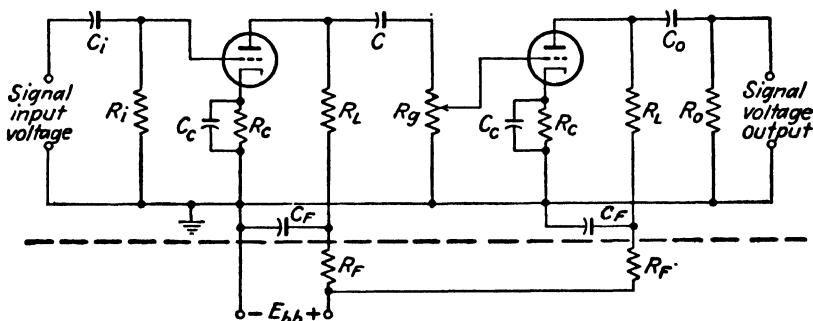



FIG. 164.—A resistance-coupled audio-frequency amplifier using triodes. All alternating components of current and voltage should be kept above the broken line.

From the tube manual it is seen that the type 6P5G tube has an amplification factor of 13.8, and it is decided to use this tube. The other data for the tube are plate resistance $r_p = 12{,}000$ ohms, and direct plate current $I_b = 2.5$ milliamperes, all values being given with the grid bias $E_c = -5$ volts, the plate voltage $E_b = +100$ volts, and the heater voltage $E_f = 6$ volts. The grid voltage is from grid to cathode, and the plate voltage is from plate to cathode.

Step 2. Direct-current Aspects. The circuit selected is shown in Fig. 164. From Step 1, $R_L = 5r_p = 5 \times 12{,}000 = 60{,}000$ ohms. Since the plate current is 2.5 milliamperes, or 0.0025 ampere, the I_bR_L direct voltage drop across these resistors will be $E_L = 0.0025 \times 60{,}000 = 150$ volts. Resistors R_F are decoupling resistors. Their purpose is to work in conjunction with capacitors C_F to form a resistance-capacitance filter (page 249). This circuit prevents "motor-boating" (page 285). Typical values for R_F are 50,000 ohms, although their value may depend on the type of power supply. The 2.5 milliampere plate current flows through these, and the direct voltage drop will be $E_F = I_bR_F = 0.0025 \times 50{,}000 = 125$ volts. Thus the power-supply voltage must be $E_{bb} = E_b + E_L + E_F = 100 + 150 + 125 = 375$ volts. Voltage E_b is the direct plate-to-cathode potential mentioned in Step 1.

Self-bias. The grid-bias voltage of -5 volts is obtained by the self-biasing arrangement of R_c and C_c. The direct plate current of 2.5 milliamperes, or 0.0025 amperes, flows (conventional direction) from plate to cathode. Thus it will flow down through cathode resistor R_c and will cause a voltage drop across R_c. This will make the upper or cathode end of R_c positive and the lower or grid end negative. To obtain a bias of -5 volts, the resistors R_c must have resistances of $R_c = E_c/I_b = 5/0.0025 = 2000$ ohms. Strictly speaking the power-supply voltage should be 380 volts to include this 5 volts, but this is of little consequence.

Step 3. Alternating-current Aspects. Starting at the left, the input condenser C_i is to prevent any direct voltage in the signal source from affecting the grid of the first tube. Since this is a voltage amplifier, C_i should be a paper condenser of perhaps 0.1 microfarad capacitance, and R_i should be about 500,000 ohms; then they will take but little current from the source of signal voltage, and will have little effect on the frequency characteristics of the circuit. It is of interest to note that the source forces signal current through this C_i-R_i combination, and that the voltage reaching the grid is the drop across R_i. Condenser C_i should be able to stand a direct voltage of, say, at least 200 volts, or external potentials may break it down.

Cathode By-pass Condenser.—Condensers C_c should be electrolytic capacitors, or other types of very high capacitance, of the order of 25 microfarads, and capable of standing about 25 volts. They are for the purpose of by-passing the alternating signal current to keep it from passing through resistor R_c. When the amplifier is passing signal, both direct and alternating components of current will be flowing from plate to cathode. If the alternating component flows through R_c, an alternating voltage will exist across it. This will combine vectorially with the alternating signal voltage existing across R_i, and feedback (page 364) will result. But if C_c is a large capacitor, which, even at the lowest frequencies, has low reactance compared with the resistance of R_c, then no appreciable alternating voltage can exist across R_c, and feedback is prevented.

Output Circuit.—Condenser C_o is the output capacitor, and prevents the direct voltage on the plate of the second tube from affecting the device connected to the output terminals. A 0.1-microfarad 600-volt paper condenser will be satisfactory. The resistor R_o is the output resistor, and is necessary only if C_o is used. A resistor of 500,000 ohms should be satisfactory, since this is a *voltage* amplifier. The input circuit C_i and R_i, and the output circuit C_o and R_o, will have negligible effect on the gain and the frequency response over the range to be passed if they are of the values listed. They can, however, affect the response materially. In a sense these are auxiliaries, and the amplifier will serve for many purposes without them.

Decoupling Network. Condensers C_f and resistors R_f form a decoupling network to prevent a form of low-frequency oscillations called "motor-boating," because of the putt-putt noise it creates. Both tubes are fed from the same power supply, and even if the power supply does contain large condensers, at very low frequencies these have appreciable reactance.

Now if a large signal is being passed, then the plate current of the second tube will contain comparatively large current variations. If the alternating signal currents flow through a power supply having appreciable impedance, then alternating signal voltages will be developed across the impedance of the power supply. This impedance is common to the plates of both tubes. If the output of the second tube causes fluctuations in the voltage to the plate of the first tube, it will also cause the plate current of the first tube to vary. This will introduce a signal into the grid of the second tube, and in some amplifiers *oscillations may result*, causing motor-boating. The oscillations are of very low frequency because they can exist only at very low frequencies at which the impedance of the power-supply filter condensers is appreciable. Also, because the grid bias is supplied by the plate current, feedback through the common impedances may act through the grid circuit of the first tube. The purposes of R_F and C_F are to keep all alternating voltages and currents above the dotted line. The resistors R_F offer opposition to current flow, and capacitors C_F offer a good path from plate to cathode for the alternating-current signal component of the plate circuit. As explained in Step 2, the values of R_F might be 50,000 ohms. The capacitors C_F often are electrolytic condensers rated at 8 microfarads and 450 volts. If the amplifier is powered with batteries, and if they are not allowed to deteriorate, then the decoupling networks may be omitted, because good batteries have negligible internal resistance. Also, for a two-stage amplifier it may be that only one decoupling network is necessary and some amplifiers will operate with no decoupling networks. This can be determined experimentally. Two-stage amplifiers are not likely to motor-boat because of the 180° phase shift in a single stage. They will motor-boat, however, especially if they are of high gain. Three-stage amplifiers are very likely to motor-boat.

Selection of C, R_L, and R_g. It is now necessary to determine C, R_L, and R_g and to study the over-all gain and frequency response. In Step 2 it was shown that R_L would be 60,000 ohms. It is common to make R_g about 500,000 ohms. In this case it will be a voltage divider so that the output can be continuously variable. This could be placed in the input of the first tube instead of R_1. It was not done because it is inadvisable to have variable contacts in low-level circuits, such as at the input of an amplifier. In general, it is better to have some amplification first. On the other hand, if large signals are to be encountered ever, so that the first tube might be overloaded, it might be better to have the voltage divider at R_1. A capacitor of 0.02 microfarad is a reasonable value for condenser C; it should have a voltage rating of perhaps 600 volts, and it *must have very high leakage resistance.* If it does not, then direct current will leak from the plate of the first tube through R_g and will affect the bias of the second tube, causing it to operate incorrectly. Often mica condensers are used for capacitor C, but very good paper condensers operate satisfactorily.

Intermediate-frequency Gain.—The gain at intermediate frequencies will be calculated as explained on page 277. The value of R_L for Eq. (80) will be the value of R_L and R_g in parallel, which for the amplifier of Fig. 164 is

(page 82) $R_L R_g/(R_L + R_g)$, or $60{,}000 \times 500{,}000/(60{,}000 + 500{,}000) = 53{,}600$ ohms. Then, using Eq. (80), $A_v = \mu R_L/(r_p + R_L) = 13.8 \times 53{,}600/(12{,}000 + 53{,}600) = 11.25$. Since a gain of only 10 per stage is needed, this is adequate.

Low-frequency Gain.—The gain at 50 cycles will now be checked. Using Eq. (83), the low frequency at which the gain is down to 70 per cent of the value at intermediate frequencies and the phase shift is 45° occurs at

$$\frac{60{,}000 + 12{,}000}{6.28 \times 0.02 \times 10^{-6}(12{,}000 \times 500{,}000 + 60{,}000 \times 500{,}000 + 12{,}000 \times 60{,}000)} = 15.5 \text{ cycles.}$$

This calculation indicates that the low-frequency response would be adequate. In fact, a coupling capacitor of $C = 0.01$ microfarad would be adequate, and probably should be used. There is no object in making the low-frequency response better than necessary; the better the low-frequency response, the greater the tendency to motor-boat. If paper condensers are used, the smaller the capacitance, the less is the area of the dielectric, and the greater should be the leakage resistance.

High-frequency Gain.—The gain at 10,000 cycles will be checked now. Using Eq. (85), the high frequency at which the gain is down to 70 per cent of the value at intermediate frequencies and the phase shift is 45° is

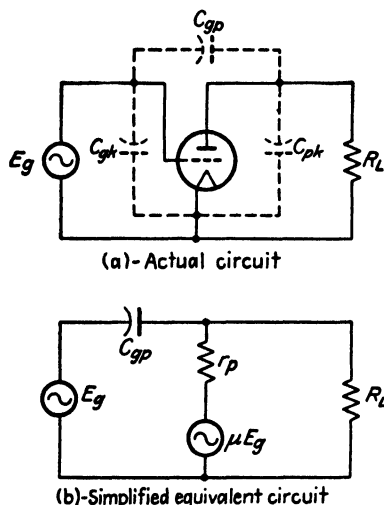
$$\frac{12{,}000 \times 500{,}000 + 60{,}000 \times 500{,}000 + 12{,}000 \times 60{,}000}{6.28 \times 75 \times 10^{-12} \times 60{,}000 \times 500{,}000 \times 12{,}000} = 216{,}000 \text{ cycles.}$$

This is an astonishingly high value, and of course is theoretical. It does indicate that the amplifier certainly is satisfactory at high frequencies. A very wide frequency response is typical of resistance-coupled amplifiers using *triodes*. The reason for making C_g' of Eq. (85) equal to 75×10^{-12} farad will be explained in the following section. The high-frequency response of the amplifier can be reduced to a lower value by connecting a small condenser in parallel with R_g of Fig. 164. Attention is called to the fact that all calculations were made at full gain, that is, with volume control R_g on the maximum setting.

Input Impedance of Vacuum Tubes.—The grid of a tube used as a voltage amplifier is biased negatively. In the example just given, the bias is −5 volts. If the peak value of the alternating signal voltage is always less than 5 volts, the grid always remains negative with respect to the cathode. Because a negative grid will not attract negative electrons, no direct current should flow in the grid circuit. This would be true if it were not for the fact that there are some positive ions in a tube. The high-vacuum tube is evacuated to the extent that the characteristics essentially are unaffected by the residual gas. But, some gas remains, some ionization occurs, a few positive ions are produced, and these constitute a very small positive current flow to the negative grid. As a result,

in the external grid circuit, conventional current flows from grid to cathode through the grid resistor, and this causes a voltage drop that *opposes* the grid-bias voltage, tending to reduce the bias actually existing between grid and cathode.

This is one reason that the grid resistor usually is 250,000 or 500,000 ohms, and seldom is over 1,000,000 ohms. Other reasons are that if the grid is momentarily driven positive by a signal overload, the grid may attract sufficient negative electrons so that it becomes highly negative and "blocks" the flow of plate current. If the grid resistor is large, it will take a long time to discharge the grid, and return it to its normal bias value. This effect is increased by large condensers at C_i and C in Fig. 164, because these condensers may take a considerable charge when a grid is driven positive, and the length of time taken to discharge these condensers through the grid resistors may be great.

FIG. 165.—Circuits for studying the effect of the interelectrode capacitances on the input impedance of a triode used in a resistance-coupled amplifier.

Although often it is quite satisfactory to assume that no current flows in the grid circuit (and that the input impedance to the grid of a tube is, accordingly, infinite), there are times when this assumption should not be made. The direct-current case was considered in the preceding paragraphs. With an alternating signal voltage applied to the grid, a similar situation exists. There are times when the grid-input impedance to alternating currents may be considered infinite, and times when it may not, depending on the type of tube, the frequency, and other factors.

Grid Input Impedance of Triodes.—Capacitance exists between the electrodes and the connecting wires *within* a vacuum tube. For a triode these capacitances are as indicated in Fig. 165*a*, and are the grid-cathode capacitance C_{gk}, the grid-plate capacitance C_{gp}, and the plate-cathode capacitance C_{pk}. Although these are *within* the tube, it is convenient to show them as indicated. The

values of these capacitances are a few micromicrofarads, as a vacuum-tube manual will indicate.

When the tube is acting as a resistance-coupled amplifier, the equivalent circuit becomes that of Fig. 165*b*. In this C_{gk} and C_{pk} are omitted. The effect of C_{gk} will be considered later; that of C_{pk} will be neglected because it is in parallel with the load resistor R_L, which for a triode is of low value.

The magnitude of the input impedance of the grid circuit of a triode is equal to the impressed alternating grid voltage divided by the alternating current that flows into the grid circuit. The angle of the input impedance will be the phase angle between the applied alternating grid voltage and the alternating grid current that flows. A somewhat advanced study, which will not be included, shows that the amplified voltage μE_g, being in series with E_g, affects the current that flows in the grid circuit and through the grid-plate interelectrode capacitance C_{gp}.

With a *pure resistance load*, the phase relations are such that the amplified voltage *increases* the 90° leading current that flows in the grid circuit, and that the apparent capacitance caused by this effect is $C_{gp}(1 + A_v)$, where A_v is the voltage gain of the stage. To find the total effective grid input capacitance, the value C_{gk} (which was omitted in Fig. 165*b*) must be added. Thus the effective input capacitance of a triode *operating as a resistance-coupled voltage amplifier* is

$$C_g = C_{gk} + C_{gp}\,(1 + A_v). \qquad (86)$$

To this must be added the wiring capacitance of the grid circuit, giving C_g', used in the amplifier calculations, Eqs. (84) and (85). For the type 6P5G with a close-fitting metal tube shield around it, $C_{gp} = 2.2$, $C_{gk} = 3.4$, and $C_{pk} = 5.5$, all values in micromicrofarads. For a gain of 11.25, as in the example given (page 287), the value of C_g' would be $C_g' = 3.4 + 2.2\,(1 + 11.25)$ plus the wiring capacitance, or 30.4 micromicrofarads plus the grid wiring capacitance. Since 75 micromicrofarads were used in the calculations just referred to, this would mean that 45 micromicrofarads existed in the wiring capacitances, a value probably to high for most amplifiers. All this means is that the theoretical high-frequency gain would be even higher than was computed. For audio frequencies, and at all but very high radio frequencies, the input capacitance may be assumed to be pure capacitance when the load is resistance.

With a load that is *resistance and capacitance,* the grid input impedance becomes resistance and capacitance, instead of approximately pure capacitance as with a load of resistance only.

With a load that is *inductance and resistance,* the grid input impedance becomes, under certain conditions, *negative* resistance and capacitance. Now an ordinary resistance is a positive resistance and absorbs power, but a negative resistance *delivers* power. This means that a tube with a load that is inductive tends to deliver power back to the grid circuit, *and this may cause oscillations* (page 356).

At very high radio frequencies the effect of **transit time** enters into determining the grid input impedance. By transit time is meant the time required for the electrons to flow from the cathode to the plate. At audio frequencies, and at radio frequencies up to perhaps 100 megacycles, transit time is of negligible importance. At very high radio frequencies the alternating electrode potentials may pass through a considerable portion of a cycle while an electron is passing across the tube. For certain conditions the electrons may arrive at the plate with such phase relations that the grid input impedance contains a resistance component that absorbs power from the alternating grid-driving signal source. This is called **electron loading.** It is of much importance at very high radio frequencies, and tubes in which the transit time is low are specially designed for such service.

Grid Input Impedance of Tetrodes and Pentodes.—When tetrodes or pentodes are used in amplifiers, the value of C'_g, which appears in Eqs. (84) and (85), is much less than with triodes. This is because the screen grid in the tetrode, and the screen and suppressor grids in the pentode, shield the control grid from the plate. Hence, Eq. (86) and the accompanying discussion do not apply to tetrodes and pentodes. The value of C'_g is, accordingly, the wiring capacitance plus what is listed in tube manuals as the input capacitance, and it is about 5 micromicrofarads only for a typical voltage-amplifying pentode. Over the usual operating range, up to perhaps 100 megacycles, there is little feedback and the input impedance is a low value of capacitance.

Resistance-coupled Amplifiers Using Tetrodes and Pentodes.—The curve of Fig. 163 and the accompanying discussion also apply to tetrodes and pentodes when used in resistance-coupled amplifiers. Just as for triodes, *theoretically* the load resistance should be

much higher than the plate resistance to utilize as much as possible of the theoretically available μE_g. It will be recalled that this amplified signal voltage must force the alternating signal current through the internal plate resistance of the tube in series with the external load resistance. If the load resistance is large compared with the plate resistance, then there will be but little drop in voltage within the tube, and, as previously mentioned, most of the amplified signal voltage μE_g will appear across the load resistance and can be utilized.

When an amplifier using tetrodes or pentodes is examined from the direct-current standpoint, it is evident immediately that the load resistance R_L must be of the same order of magnitude as, or considerably less than, the plate resistance. The reason for this is that the plate resistance of tetrodes and pentodes is very high, of the order of from 500,000 to 1,500,000 ohms. Suppose that a pentode has a plate resistance of 1,000,000 ohms, that the direct plate current is only 1.0 milliampere, and that an *attempt* is made to make the load resistance five times the plate resistance as might be done with a triode. Then the load resistance R_L would be 5,000,000 ohms, and with 1.0 milliampere through it the direct voltage drop would be 5000 volts. Thus the power supply would have to be greater than 5000 volts, which is unreasonable for equipment of this nature. Because of this fact, when tetrodes are used, the load resistance is about the same as the plate resistance, and when pentodes are used, the load resistance is much less than the plate resistance.

When tetrodes are used, care must be taken to ensure that the plate potential never falls to a value such that operation is in the erratic plate-current region of Fig. 104*b*, page 189. This often is done by having the direct screen-grid potential at less than 100 volts, and the direct plate potential at several hundred volts. Of course the instantaneous plate potential increases and decreases when the tube is passing a sighal, and the minimum plate potential should be greater than the screen potential. With pentodes, the suppressor grid prevents secondary emission from affecting the tube characteristics, and, hence, these precautions need not be observed. The screen grid and the plate may, accordingly, be operated at about the same direct voltages with respect to the cathode. Pentodes are used more widely than tetrodes in modern amplifiers, and in the following example pentodes will be employed.

Illustrative Problem.—It is desired to design an audio-frequency voltage amplifier using pentodes which will have a variable voltage gain from zero up to a minimum of 80 decibels and which will be substantially flat from 50 to 10,000 cycles.

Solution.—Step 1. General Aspects. A voltage gain of 80 decibels is an amplification ratio of [from Eq. (28), page 96] $E_1/E_2 = 10^{0.05 \times n} = 10^{0.05 \times 80} = 10{,}000$. This will require two tubes each producing a voltage amplification of 40 decibels and a voltage gain of 100. This gain is readily obtainable with a voltage-amplifying pentode. If the plate resistance of the selected tube is 1,000,000 ohms, if it has an amplification factor of 1000, and if a single tube with a load resistance of 250,000 ohms is used, the gain of such a stage from Eq. (80), page 273, will be $A_v = \mu R_L/(r_p + R_L) = 1000 \times 250{,}000/(1{,}000{,}000 + 250{,}000) = 200$, which will be far more than adequate.

It is decided to use a 6J7, which is a triple-grid tube which may be connected externally as a voltage-amplifying suppressor-grid tube. From the tube manual it is found that this tube has a plate resistance of 1.0 megohm, a grid-plate transconductance of 1185 micromhos, and an amplification factor of $\mu = r_p \times g_m = 1.0 \times 10^6 \times 1185 \times 10^{-6} = 1185$. These data hold when the direct voltage between plate and cathode is 100 volts, when the screen-grid voltage is 100 volts, the control-grid voltage is -3 volts, and the suppressor grid is connected to the cathode. Under these conditions the direct plate current will be 2 milliamperes. An examination of the characteristic curves given in the manual show that all data except the plate current will be essentially the same if the tube is operated at -4.5 volts between control grid and cathode instead of -3 volts. The direct plate current will be reduced to about 0.5 milliampere, which is very helpful as will be seen later. The screen-grid current with 100 volts on the screen and -3 volts on the control grid is 0.5 milliampere. Screen-grid curves for a control-grid bias of -4.5 volts are not given in the manual, and it is estimated that at a control-grid voltage of -4.5 volts the screen-grid current will be 0.25 milliampere. The heater voltage is 6.3 volts, and it will be obtained from filament-heating windings on the transformer used in the power supply.

Step 2. Direct-current Aspects. The circuit selected is shown in Fig. 166. From Step 1, the load resistors R_L can be about 250,000 ohms, and the current through these resistors will be 0.5 milliampere, giving a direct voltage drop across R_L of $250{,}000 \times 0.0005 = 125$ volts. Since the screen-grid voltage is to be 100 volts, and since the screen grids and the plates are to be connected together as indicated, the direct voltage drop across R_d must be 125 volts, the same as across R_L. The value of R_d will be $R_d = E/I = 125/0.00025 = 500{,}000$ ohms. As explained in Step 1, the screen-grid current was estimated because it was not given in the tube manual for a control-grid voltage of -4.5 volts. Thus 500,000 ohms is an approximation, and perhaps some measurements of screen-grid currents and voltages should be made to determine a more exact value. Such matters usually are not critical.

Resistors R_F are decoupling resistors used in conjunction with capacitors

C_F to reduce the amount of signal-current flow in the common impedance of the power-supply filter to a minimum, thus preventing oscillations called "motor-boating." As previously explained for the triode amplifier, whether or not either or both decoupling networks are used will depend on conditions; often they are not necessary with two-stage amplifiers. In any event, both the plate and screen currents flow through them, giving a total current of $0.5 + 0.25 = 0.75$ milliamperes. If these decoupling resistors each are 100,000 ohms, then the direct voltage drop across them will be $0.75 \times 100{,}000 = 75$ volts, and the voltage E_{bb} of the power supply must be $E_{bb} = E_b + E_L + E_F = 100 + 125 + 75 = 300$ volts. If the usual power-supply components are used, a direct voltage of about 300 volts will be available. If such a power supply is employed, then the decoupling filter may as well be used to reduce the voltage to the desired amount and to act as a precaution against motor-boating.

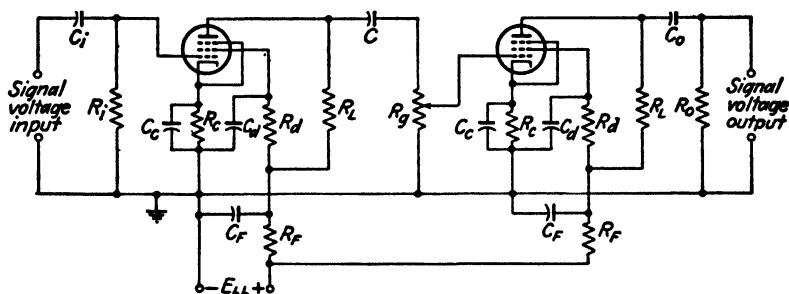



FIG. 166.—A resistance-coupled audio-frequency voltage amplifier using pentodes.

The grid-bias voltage of -4.5 volts will be obtained by the self-biasing arrangement of Fig. 166. Since the combined plate and screen currents of 0.75 milliampere flow through R_c, its value must be $R_c = E/I = 4.5/0.00075 = 6000$ ohms.

Step 3. Alternating-current Aspects. The input capacitor C_i, the output capacitor C_o, the cathode capacitors C_c, and the decoupling capacitors C_F were all discussed on page 285 for the triode amplifier. In general, the same statements apply to the use of these capacitors in the pentode amplifier of Fig. 166. The screen-grid resistor R_d, placed in the circuit to reduce the power-supply voltage to the correct value for impressing on the screen grid, must be by-passed to ground by a condenser C_d. The reason for this is that the screen grid is placed in the tube to shield the control grid from the plate. Of course the screen grid must have a positive voltage on it, or electrons will not pass through it to the plate. But, in an amplifier, the potential of the screen grid must not vary, otherwise the screen grid will cause variations in the current flow, and in amplification it is desired that the control grid only have an effect on the alternating plate-current flow. Thus condenser C_d is connected from the screen grid direct to ground. From a direct-current standpoint, this condenser charges and discharges if for any reason the screen-grid potential attempts

to change, and thus condenser C_d tends to hold the screen-grid potential constant. From an alternating-current standpoint, condenser C_d offers a path of low impedance from screen grid to ground and negligible alternating voltage can exist between screen grid and ground. The value of C_d is not critical, although it does affect the signal response of the circuit to some extent. A paper condenser of about 0.1 microfarad and capable of withstanding the power-supply voltage often is used in a circuit such as Fig. 166. A coupling capacitor of $C = 0.005$ microfarad is a common value for such amplifiers and will be used. It should have a voltage rating of about that of the power supply.

The gain at intermediate frequencies will be determined by Eq. (80) and modified as explained on page 286. The equivalent resistance of R_L and R_g in parallel is $R_e = R_L R_g/(R_L + R_g) = 250{,}000 \times 500{,}000/(250{,}000 + 500{,}000) = 167{,}000$ ohms. Then, using Eq. (80), $A_v = \mu R_L/(r_p + R_L) = 1185 \times 167{,}000/(1{,}000{,}000 + 167{,}000) = 170$. Now this calculation shows that each stage of Fig. 166 will produce a voltage gain *far in excess* of the gain of 100 that each stage must furnish to meet the requirements. The question arises: Why not reduce the values of R_L and R_g, and also perhaps use the bias voltage at -3 volts instead of -4.5 volts? The answer is that this would be a very reasonable thing to do. However, the calculations just made were *entirely theoretical*, and experience shows that such calculations usually are high. Also, if the gain actually reaches 170 per stage, then the amplifier will have a total gain of 28,900, or about 90 decibels, and this is only 10 decibels greater than the *minimum* requirements. Thus it is perhaps advisable to continue with the values originally assumed.

The gain at low frequencies will be checked now. Using Eq. (83), the low frequency at which the gain is down to 70 per cent of the value at intermediate frequencies and the phase shift is 45° occurs at

$$\frac{250{,}000 + 1{,}000{,}000}{6.28 \times 0.005 \times 10^{-6}(1{,}000{,}000 \times 500{,}000 + 250{,}000 \times 500{,}000 + 1{,}000{,}000 \times 250{,}000)} = 45.5 \text{ cycles.}$$

This calculation indicates that the low-frequency response will be adequate; that is, the amplifier readily passes frequencies below 50 cycles.

The gain at high frequencies will be checked now, using Eq. (85). This equation calls for C_g', the total grid input capacitance, including the capacitance of the wiring and the tube. As explained on page 290, this will be lower than with triodes. A typical value is 20 micromicrofarads, 15 for the wiring, and 5 for the tube input capacitance. Using Eq. (85), the upper or high frequency at which the gain is down to about 70 per cent of the value at intermediate frequencies and the phase shift is 45° occurs at

$$\frac{1{,}000{,}000 \times 500{,}000 + 250{,}000 \times 500{,}000 + 1{,}000{,}000 \times 250{,}000}{6.28 \times 20 \times 10^{-12} \times 250{,}000 \times 500{,}000 \times 1{,}000{,}000} = 55{,}500 \text{ cycles.}$$

This calculation indicates that the high-frequency response will be adequate; that is, the amplifier readily passes frequencies above 10,000 cycles.

It will be noted that the theoretical band width of the typical *triode* resistance-coupled amplifier (page 287) is much wider than the theoretical band width of the *pentode* resistance-coupled ampli-

fier just considered. In general this is true: Triode resistance-coupled amplifiers have lower gain and wider band width than do pentode resistance-coupled amplifiers. Triode amplifiers are not satisfactory for voltage amplifiers at radio frequencies because of the feedback between plate and grid and the resulting tendency to oscillate (page 356).

Alternate Method of Computing Gain of Pentode Amplifiers.—In the preceding pages both triode and pentode amplifiers were

This illustration shows a complete interstage coupling circuit for a resistance-coupled voltage amplifier using pentodes. This unit, called by the trade name "Couplate," contains resistor R_L (250,000 ohms), resistor R_g (500,000 ohms), and capacitor C (0.01 microfarad) of Fig. 166. In addition, it contains a 250-micro-microfarad capacitor connected as a radio-frequency by-pass condenser across R_L of Fig. 166 to ground. (*Centralab Division of Globe-Union, Inc.*)

considered to be generators of open-circuit voltage μE_g and of internal resistance r_p. The voltage was assumed to force the current through the internal plate resistance r_p and the external load resistance R_L in series. This is the common way of considering electric circuits. For pentodes, an alternate method of regarding the tube and load, and of making calculations, is available, as will be explained now.

The basic equation for calculating the gain of an amplifier is Eq. (80)

$$A_v = \frac{\mu R_L}{r_p + R_L}.$$

For the practical pentode resistance-coupled amplifier, r_p is many times larger than R_L; for instance, in the example of the preceding section $r_p = 4R_L$, and, in such amplifiers, r_p often is 6 to 10 times

R_L. If it is assumed that R_L is negligible with respect to r_p, then adding R_L to r_p in the denominator will produce negligible change in the denominator, so that the equation becomes

$$A_v = \frac{\mu R_L}{r_p} = g_m R_L, \qquad (87)$$

where $g_m = \mu/r_p$, as explained on page 187.

With this simple equation the amplification per stage can be calculated with a fair degree of approximation. For example, if $\mu = 1185$, $R_L = 250{,}000$ ohms, $r_p = 1{,}000{,}000$ ohms, and $g_m = 1185$ micromhos as for the pentode amplifier of the preceding section, then the two methods give for the gain per stage

$$A_v = \frac{\mu R_L}{r_p + R_L} = \frac{1185 \times 250{,}000}{1{,}000{,}000 + 250{,}000} = 237, \text{ or } 47.5 \text{ decibels, and}$$

$$A_v = g_m R_L = 1185 \times 10^{-6} \times 250{,}000 = 296, \text{ or } 49.5 \text{ decibels.}$$

Of course, there is quite a difference between a gain of 237 and 296, but when expressed in decibels the difference is minor, and it must be remembered that in radio work amplification often is expressed in decibels. As a further justification of this alternate method, it must be remembered that the data given in tube manuals are for an "average" tube and that a specific tube may differ considerably, and also that resistors, etc., differ quite widely from the marked values. Furthermore, amplifiers usually contain volume controls. Therefore, all factors considered, Eq. (87) is a simple way of calculating the approximate gain per stage. Of course it does not give the width of the band passed, but often such data are not necessary. Often it is known from previous experience that if certain values of resistance, etc., are used that the band width will be satisfactory. It is very important to note that Eq. (87) *should not be used with triodes* because in such amplifiers R_L is not negligible with respect to r_p.

Thévenin's and Norton's Theorems.—The method of calculation, using Eq. (87), is in fact a method of basic importance used in communication. The usual way of considering circuits is to use Thévenin's theorem, often called the Ohm's law method. This considers a circuit containing a vacuum-tube amplifier as a *series* circuit containing a source of *constant voltage.* The alternate method, often used in radio, is to consider a vacuum-tube amplifier as a *parallel* circuit containing a source of *constant current.* These

two methods will be considered now. Of course the two methods apply to circuits containing devices other than vacuum tubes.

For the familiar circuit of Fig. 167*a* the amplified output voltage E_L is given by Eq. (79),

$$E_L = \frac{\mu E_g R_L}{r_p + R_L}.$$

In accordance with the rules of algebra, it is permissible to multiply *both* the numerator and denominator by the same term, and multiplying by r_p gives

$$E_L = \frac{\mu E_g r_p R_L}{r_p(r_p + R_L)} = \left(\frac{\mu}{r_p}\right)(E_g)\left(\frac{r_p R_L}{r_p + R_L}\right)$$
$$= (g_m E_g)\left(\frac{r_p R_L}{r_p + R_L}\right), \qquad (88)$$

where the mutual conductance g_m equals μ/r_p.

FIG. 167.—Circuit for explaining Thévenin's theorem (*a*), and circuit for explaining Norton's theorem (*b*).

As indicated, Eq. (88) consists of two parts. The first term is $g_m E_g$, the product of the mutual conductance of the tube and the alternating signal voltage impressed on the grid of the tube. The product of conductance and voltage is current, and thus, $g_m E_g = I_{psc}$, the alternating signal plate current *with no load* in the plate circuit, that is, with the plate circuit short-circuited. This is true because the mutual conductance g_m is measured with no load in the plate circuit (page 186). The second term is $r_p R_L/(r_p + R_L)$, and from page 82, this is the parallel impedance of r_p and R_L. From this it follows that the two circuits of Fig. 167 are equivalent, because with the same applied alternating signal voltage E_g, the same amplified output signal voltage E_L is obtained. On the basis of such reasoning, somewhat more generalized of course, Thévenin's and Norton's theorems can be stated as follows:

Thévenin's Theorem.—The current that flows through an impedance Z_L connected to the two terminals of a generator equals the open-circuit voltage E_{oc} at the generator terminals divided by the sum of the internal generator impedance and the load impedance. That is,

$$I_L = \frac{E_{oc}}{(Z_g + Z_L)}. \tag{89}$$

Such a circuit, illustrated with a vacuum tube as an amplifier, appears in Fig. 167*a*.

Norton's Theorem.—The current I_L that flows through an impedance Z_L connected to the two terminals of a generator of internal impedance Z_g is the same as if the impedance Z_L with Z_g in parallel is connected to a constant-current generator whose current is the same as the current that flows through the generator terminals when they are short-circuited. That is,

$$I_L = \frac{E_L}{Z_L} = \frac{I_{sc}Z_g Z_L}{(Z_g + Z_L)} \times \frac{1}{Z_L} = \frac{I_{sc} Z_g}{(Z_g + Z_L)}. \tag{90}$$

Such a circuit, illustrated with a vacuum tube as an amplifier, appears in Fig. 167*b*.

As a further proof of Eq. (90), using ordinary series-circuit theory, the current I_{sc} that flows when a generator of open-circuit voltage E_{oc} and internal impedance Z_g is short-circuited is $I_{sc} = E_{oc}/Z_g$, and $E_{oc} = I_{sc}Z_g$. When this value is substituted in Eq. (89) for Thévenin's theorem, Eq. (90) for Norton's theorem results. These two theorems and Eqs. (89) and (90) may be used for any circuits, because they are interchangeable. However, Thévenin's equation is in general use in the form of a generalized Ohm's law. There is a growing tendency to use Norton's theorem in radio, particularly in high-impedance circuits.

Further Comments Regarding Amplifier Design.—The design of audio-frequency resistance-coupled voltage amplifiers has been considered in much detail in the preceding pages. As was previously mentioned, this was because these amplifiers have wide use in practice, and because once their theory is understood, it is relatively easy to grasp the basic principles of voltage amplifiers of other types, for example, those used at radio frequencies.

From the purely practical standpoint this detailed discussion is not justified, because amplifiers can be "designed" and constructed from data given in radio manuals and handbooks. These sources give complete information, such as the values of resistors, capacitors, voltages to be used with various tubes, and the voltage gains that result.

Compensated Amplifiers.—As has been explained, the usual resistance-coupled amplifier will pass a rather wide band of fre-

quencies. This band, however, is not sufficiently wide for television purposes, where the video amplifier must pass a band of from about 30 to 3,000,000 cycles or wider, depending on the image definition desired. These **video amplifiers** are used in television to amplify the signals from the image pickup camera and to increase the amplitude of these signals to the point at which it is possible to modulate the carrier (page 405). Also, video amplifiers are used

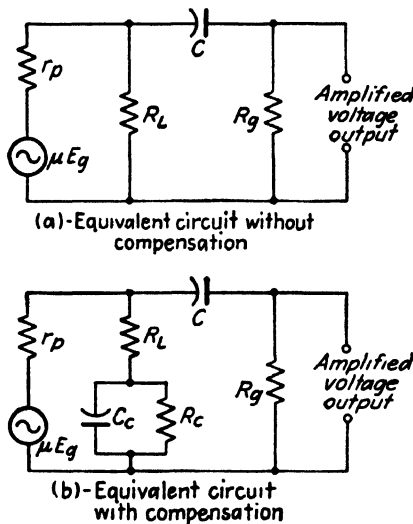



FIG. 168.—Equivalent circuits of a wide-band resistance-coupled voltage amplifier at *low* frequencies, without and with compensation.

in television receivers to increase the magnitudes of the signals until they can drive the image-reproducing cathode-ray tube. Thus the video amplifiers of a television system correspond to the audio amplifiers of a radio system.

The design of video amplifiers is a very specialized subject, and several types of amplifiers have been devised. One basic scheme is to install low-frequency and high-frequency compensation in a resistance-coupled amplifier, using pentodes. Triodes are not used because of their unsuitability as high-frequency voltage amplifiers.

Low-frequency Compensation.—The diagram of Fig. 168*a* is the equivalent circuit of a resistance-coupled voltage amplifier *at low frequencies* (page 278). The problem is to select units such that the low-frequency response will be good, down to 30 cycles or less.

This means that the amplified signal voltage μE_g should force a large current through R_g, so that the voltage drop across R_g, which is the amplified voltage output, will be large.

Since pentodes are used in video wide-band amplifiers, the plate resistance, r_p of Fig. 168 will be high, and the current flowing through r_p largely will be independent of the rest of the circuit. If the resistance of R_L is high with respect to R_g, and if the reactance of the coupling condenser is as low as possible, then most of the plate signal current flowing from the tube will be forced through the grid resistor R_g and the voltage drop across R_g will be large. Since this is the amplified output voltage, the gain per stage will be good at low frequencies.

As will be shown later, there are reasons why the load resistor R_L must be low, although *from the standpoint of the low-frequency response this is undesired.* Also, there are practical reasons why the capacitance of the coupling capacitor C cannot be made as high (and the reactance as low) as desired. For these reasons, the compensating network C_c-R_c of Fig. 168*b* is used. The units selected are such that at the low-frequency values of, say, 30 cycles, the reactance of the condenser C_c is so great that its effect in the circuit is negligible; then, the effective load resistance is approximately $R_L + R_c$. Because of this increase in the effective load resistance, the low-frequency response will be maintained as explained in the preceding paragraph. At intermediate and high frequencies, however, the reactance of the compensating condenser C_c is negligible, from an alternating-current standpoint R_c is short-circuited, and the effective load resistance is R_L.

High-frequency Compensation.—The equivalent circuit of the basic resistance-coupled amplifier at high frequencies is shown in Fig. 169*a*. Although every effort is made to minimize the value of the tube input capacitance and the grid wiring capacitance, C'_o, the effect of this combined capacitance becomes very important at the high video frequencies of about 3,000,000 cycles.

As was implied under low-frequency compensation, for a wide response the coupling resistor R_L should be low. A simple explanation for this is as follows: If R_L is low, then the nature of the circuit to the right of R_L will have little effect on the frequency response. In other words, if R_L is low, the gain will be low at all frequencies, the effect of a parallel capacitance such as C'_g will be

less than if R_L is high, and the gain will be good over a wide band. For this reason, even when pentodes having plate resistances of the order of a million ohms are used, the values of R_L are of the order of a few thousand ohms.

The gain at high frequencies can be increased by inserting a small inductor L in series with the load resistance R_L, as shown in Fig. 169*b*. In this circuit R_g is omitted because it is so high in

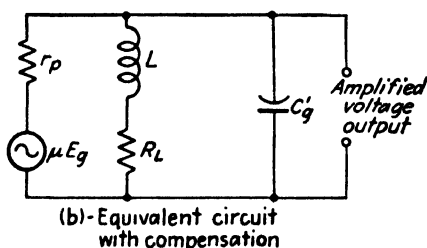
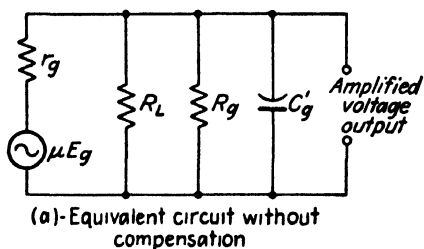



FIG. 169.—Equivalent circuits of a wide-band resistance-coupled voltage amplifier at *high* frequencies, without and with compensation.

comparison with R_L. Thus in Fig. 169*b* the parallel branches comprise a parallel resonant circuit, and this parallel resonant circuit now constitutes the load on the vacuum tube. Also, whatever voltage exists across the parallel resonant circuit is the amplified output voltage. The circuit is made resonant in the high-frequency region where the output of the uncompensated amplifier tends to decrease. In this region the input impedance of the parallel combination rises (page 83), and since the circuit may be regarded as a constant-current circuit (as explained under low-frequency compensation), the voltage output rises or at least is maintained instead of falling off. Here again the effect of R_L enters. From the theory of parallel circuits, if R_L is high, the Q will be very low and the resonance characteristics will be spread

over a wide frequency range. But if R_L is low, the Q will be high, and the effect of the parallel resonant circuit can be concentrated in the high-frequency region as desired.

Features of Compensated Amplifiers.—Special tubes are available for television video (wide-band) compensated amplifiers. In general, high gain is sacrificed for wide frequency response, and load resistors of but a few thousand ohms are used. Of course the low-frequency compensation and the high-frequency compenation would be combined into one circuit, such as Fig. 170. Not only must the response be wide but also the delay distortion must be low.

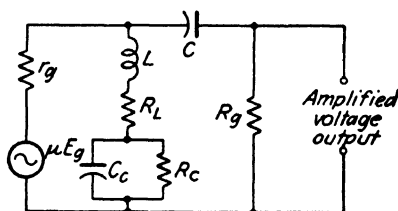



FIG. 170.—Showing how the low-frequency and high-frequency compensation of Figs. 168 and 169 are connected in a compensated wide-band amplifier.

In amplifiers in which the frequency response is wide and substantially the same at all frequencies, the delay distortion *is* low. This is an interesting and useful principle to remember. The frequency response or gain of a resistance-coupled amplifier falls off only because of the effect of capacitive reactance, and in these amplifiers only such reactances can produce phase shift (disregarding the basic 180° shift produced by a tube working into a pure resistance load). Thus, when the gain or response of an amplifier is exactly the same at various frequencies, the capacitances are ineffective, the equivalent circuit is pure resistance (Fig. 160), and no phase shift can be produced.

Transformer-coupled Audio-frequency Voltage Amplifiers.—In introducing this subject it is well to mention that there are four types of transformers that sometimes are used in audio-frequency circuits. (*a*) There is the **power-supply transformer** discussed in the preceding chapter. (*b*) **Input transformers** are used between a microphone, or a telephone transmission line, and the grid of the first tube of the amplifier. This serves to increase the feeble signal voltage and to isolate the microphone or the line from the amplifier, thereby reducing the tendency for noise to be introduced. (*c*) The

several tubes of an audio-frequency amplifier are sometimes connected together with **interstage transformers.** These will be considered in the following paragraphs. (*d*) The output of an amplifier often is fed to the load, such as a loudspeaker, through an **output transformer.** This will be considered in Chap. IX.

The use of input, interstage, and output transformers is indicated by the audio-frequency amplifier circuit of Fig. 171. As previously mentioned, the input transformer can be regarded as a voltage

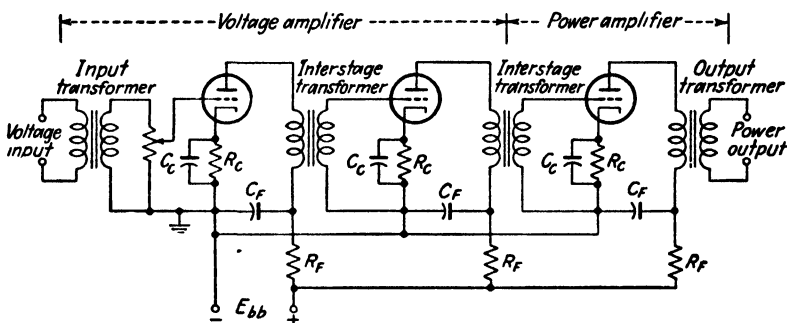



FIG. 171.—Circuit of a transformer-coupled audio-frequency amplifier. The first two tubes are voltage amplifiers and the last tube is a power-output tube. In some instances the last interstage transformer would be called a driver transformer, particularly if the grid of the last tube drew appreciable power.

transformer. It often has a high voltage step-up ratio such as 1 to 25 to increase the feeble signal voltage developed by the microphone, or other similar source, before this voltage is impressed on the grid of the first tube. The volume control indicated is of the order of 500,000 ohms, and the grid of the tube is biased negatively; thus the power handled by the input transformer is negligible. For this reason the input transformer is regarded as a voltage transformer instead of as an impedance transformer. It may be advisable to place the volume control between the first and second tubes, although it sometimes is used as indicated. Of course just any 1 to 25 ratio transformer will not do. The transformer must have the correct impedances so that it will operate correctly. For example, if the amplifier is to be fed from a microphone preamplifier that often has an output of 50 ohms, then the input transformer must be designed to work out of such an impedance.

The two interstage transformers in some amplifiers may be identical. They usually have a step-up ratio of about 1 to 3. Here

again they may be regarded as voltage transformers because the power passed to the negatively biased grids connected to the transformer secondaries is negligible. The operation of the output transformer will be discussed in the following chapter. The resistors and capacitors in the cathode circuits furnish the negative grid bias, as explained on page 285. The resistors and condensers in the plate circuit are for decoupling purposes, as explained on page 285.

Audio-frequency Transformers.—The transformers used in an amplifier such as Fig. 171 have cores of some good magnetic flux-conducting material, such as thin silicon-steel laminations or powdered compressed magnetic alloys like the Permalloys (page 63). The material used should have high permeability at the feeble signal current strengths, and it should have low hysteresis and eddy-current losses. The thin laminations, or the powdering process, reduce the eddy-current losses. The primaries of the transformers usually carry the direct-current component of the plate current, and the magnetic cores must be sufficiently large so that they do not saturate (page 245). Such transformers should be tested with the same direct-current component in the primary that will be used in practice, and with alternating signal voltages and currents of operating magnitudes. In audio-frequency transformer-coupled voltage amplifiers, phase shift is of little consequence because the ear is not sensitive to such distortion. But the ear is quite sensitive to both frequency distortion (not passing all frequencies in same relative amplitude) and nonlinear distortion (creation of harmonics). For these reasons, good transformers should be purchased. Since transformers produce some stray magnetic fields, mutual magnetic coupling exists. This is minimized by placing the transformers in cases of good magnetic flux-conducting ability (page 117). Sometimes additional shields are placed around the assembled transformers, and often they are arranged on the mounting base or chassis so that coupling is minimized. Transformers are purchased and are not assembled on the job, and care should be taken to select transformers made by a reliable manufacturer whose ratings are conservative.

Characteristics of Transformer-coupled Amplifiers.—In general, such amplifiers are just as good as the transformers used. With ordinary transformers the frequency range is fairly uniform from about 50 to 10,000 cycles, and the nonlinear distortion is not

noticeable. It is possible to purchase transformers that operate over a wider range.

As previously mentioned, the interstage transformers usually have a step-up ratio of 1 to 3. In the early days of radio the audio-frequency portion of radio-receiving sets used transformers with higher ratios, perhaps 1 to 6. The frequency response of such transformers was very poor. One reason was that with such high ratios many secondary turns, having high distributed capacitance between turns, were necessary. This high secondary capacitance, and the high magnetic flux leakage caused by the use of poor magnetic cores, combined to cause the transformer to "peak" and to give a much higher voltage ratio, or gain, at about 5000 cycles than at other frequencies. In modern transformers, good magnetic cores, lower turns ratio (usually about 1 to 3), and other improvements have eliminated to a great extent this frequency distortion. Sometimes a resistor of the order of a few hundred thousand ohms is connected across the secondary of a transformer to improve its response characteristics. This reduces the gain slightly.

In general, *interstage* transformer coupling is used only in audio-frequency voltage amplifiers *that employ triodes*, and then generally only with triodes that have plate resistances of the order of 10,000 ohms and amplification factors of about 10. Audio-frequency transformers for tubes with high plate resistances, such as certain triodes and all voltage-amplifying tetrodes and pentodes, would require an excessively large number of turns to obtain the required inductance to give the impedance necessary for operation with tubes of such high plate resistance. (Low-impedance transformers connected across tubes with high plate resistance would tend to short-circuit the tubes, and all the available amplified voltage μE_g would be lost inside the tube as a voltage drop across the plate resistance, and little would be delivered to the transformer for useful purposes.) Large numbers of turns would have high distributed capacitance, and this would result in frequency distortion.

Equivalent Circuit of Transformer-coupled Amplifier.—This is shown, for a single stage, in Fig. 172. An interstage transformer is designed with a primary that has high impedance with respect to the plate resistance of the tube from which it works. Such a primary will draw but little alternating signal current from the tube, there will be but little $I_p r_p$ voltage drop within the tube, and most of the amplified voltage μE_g will be delivered to the trans-

former primary. This will then be increased in accordance with the voltage step-up ratio of the transformer, and the output voltage across the secondary will be

$$E_O = \mu E_g N, \tag{91}$$

where μ is the amplification factor of the tube, E_g is the alternating signal voltage impressed on the grid, and N is the turns ratio of the transformer. From this equation the voltage gain per stage is

$$A_v = \frac{E_O}{E_g} = \mu N. \tag{92}$$

This equation applies only at midfrequencies, such as 1000 cycles, and only when the correct transformers, designed to work with the tubes used, are selected. The equation is quite approximate.

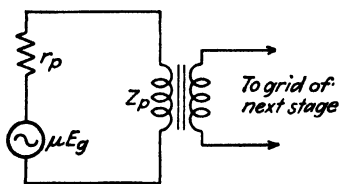



FIG. 172.—Simplified circuit for a single-stage transformer-coupled audio-frequency amplifier. The amplified output voltage is impressed between the grid and cathode of the next stage if further amplification is desired.

Because audio-frequency transformers are purchased, rather than built on the job, but little theory is needed in the design of the voltage-amplifying portion of a transformer-coupled amplifier, as the following example indicates.

Illustrative Problem.—The last tube of Fig. 171 is a power-output tube, and it requires a peak signal-driving voltage of 50 volts to produce full power output. The minimum input signal voltage available has an effective value of 0.001 volt. Select tubes and transformers that will increase this input voltage so that full power output may be obtained. The power amplifier is assumed to draw negligible signal power.

Solution.—Step 1. An effective value of 0.001 volt has a peak value (assumed sinusoidal) of $0.001 \times 1.414 = 0.001414$ volt. The voltage gain of the amplifier must be $50/0.001414 = 35{,}300$ or $n = 20 \log_{10} 35{,}300 = 91$ decibels, approximately.

Step 2. If a 1-to-25-ratio input transformer is to be used, the tubes and interstage transformers must furnish a voltage gain of $35{,}300/25 = 1410$. If two stages of voltage amplification are to be used, as in Fig. 171, and if interstage transformers having a voltage step-up ratio of 1 to 3 are to be used, then the tubes must furnish a gain of $1410/(3 \times 3) = 157$. Then, each voltage-amplifying tube must furnish a gain of $\sqrt{157} = 12.5$.

Step 3. From the tube manual it is found that types 56, 76, or 6P5-G have plate resistances of about 10,000 ohms and amplification factors of about 13.8. The type 56 is selected because it has a 2.5-volt heater, and this happens to be the same as the filament voltage needed for the power-

output tube; thus the same filament-supply winding can be used for all tubes. The theoretical gain per stage will be greater than required, but this is a good point, since the design is approximate. The theoretical gain with the use of the type 56 tubes is equal to the product of the voltage gain of the input transformer, the amplification factor of the first tube, the voltage gain of the first interstage transformer, the amplification factor of the second tube, and the voltage gain of the second interstage transformer; that is, the theoretical voltage gain is $25 \times 13.8 \times 3 \times 13.8 \times 3 = 42,800$, or about 93 decibels. Satisfactory transformers for working between the tubes used will be selected from the transformer manual of a reliable manufacturer.

Step 4. The resistors and condensers in the cathode leads were discussed on page 285. If the type 56 tubes considered here are to be operated with a plate voltage of 100 volts, the grid bias should be -5 volts, and the plate current will be about 2.5 milliamperes. The cathode resistors should be $R_c = 5/0.0025 = 2000$ ohms. The cathode by-pass condensers C_c should be electrolytic capacitors of about 25 microfarads capacitance and should be able to stand about 25 volts. The decoupling resistors R_F in the plate leads of the first two tubes should be of the order of 50,000 ohms, depending on the voltage of the power supply, and other features. In any event, if the first two tubes are to have about 100 volts on the plate, these resistors should have I_bR_F direct voltage across them that reduces the power-supply voltage to 100 volts. The decoupling capacitors C_F should be electrolytic capacitors of the order of 8 microfarads and should have a voltage rating of about 450 volts.

Radio-frequency Voltage Amplifiers.—The requirements of radio-frequency amplifiers are different from those of audio-frequency amplifiers. A good audio-frequency amplifier would amplify equally well over the entire audio band of from at least 50 to 10,000 cycles. Now a radio-frequency amplifier *does not* operate over the entire radio-frequency band. This band, or more correctly, spectrum, unlike the audio band, is extremely wide, extending from perhaps a hundred thousand cycles to many millions of cycles. It is not even desired that a single radio-frequency amplifier will amplify over this entire spectrum. In fact, much of the selectivity of a radio-receiving set, and its ability to receive one station and reject others, depends upon the fact that radio-frequency amplifiers operate only over a certain frequency band within the radio-frequency spectrum. This ability to select and amplify certain frequencies, and to reject all other frequencies, is one of the outstanding features of radio-frequency amplifiers.

The width of the band that must be passed and the position of the band in the radio spectrum vary with the type of service and other similar factors. Of importance in determining this is the type of modulation required. This subject will not be discussed

in detail until Chap. XI. But as mentioned on page 41, in radio the low-frequency speech, or code, to be transmitted is raised or translated by modulation to radio frequencies, perhaps to many millions of cycles per second.

If the discussion is confined at present to the familiar and widely used system called "amplitude modulation," then it can be stated that the signals existing at audio frequencies are moved or translated to the radio-frequency spectrum without a change in the relative magnitudes of the original signals. In this process, as used in radio-broadcast equipment, a high-frequency signal of single frequency called the **carrier** and the audio-frequency signal are fed simultaneously into a special circuit arrangement, and modulation occurs. As a result of this process of modulation, the carrier frequency and two **sidebands** exist in the output. These two sidebands are the original messages to be transmitted, but they are raised to the radio-frequency region. If an audio-frequency signal of from 50 to 10,000 cycles is used to modulate a carrier of 1,000,000 cycles, the resulting amplitude-modulated signal will consist of one sideband from 990,000 cycles to 999,950 cycles, of the carrier of 1,000,000 cycles, and of the upper sideband of 1,000,050 to 1,010,000 cycles. Thus the band width of the entire amplitude-modulated signal is now from 990,000 to 1,010,000 cycles; that is, it is 20,000 cycles wide. Again confining the discussion to an amplitude-modulation radio system, as will be done also in the pages immediately following, a radio-frequency amplifier to handle the signal just described would operate at a center frequency of 1,000,000 cycles and would amplify a band 20,000 cycles wide.

Because of these requirements, radio-frequency amplifiers employ tuned frequency-selective circuits. These usually are tuned parallel circuits, and often are, as will be seen later, in the form of **radio-frequency transformers** with tuned primaries and secondaries. The characteristics of these transformers largely determine the frequency characteristics of the radio-frequency voltage amplifiers. The dotted lines of Fig. 173 give the ideal characteristics of a radio-frequency voltage amplifier. The solid curve gives the general shape of the response curves that can be obtained in practice.

Transformer-coupled Radio-frequency Voltage Amplifiers.—It is possible to couple two stages of radio-frequency amplification with a radio-frequency transformer consisting of a primary coil

and a secondary coil with mutual inductance between them. The primary would be placed in the plate lead of one tube and the secondary would be placed in the grid circuit of the second tube. Such an arrangement would be similar to Fig. 171, except that the tubes would be tetrodes or pentodes to prevent radio-frequency feedback between the plates and control grids. Also, air cores, or cores of special powdered and compressed magnetic material having low magnetic losses at radio frequencies, would be used. Such an amplifier would be an untuned radio-frequency amplifier, and would amplify over a wide range. The word "untuned" is misleading, however, because any coil has series inductance and parallel distributed capacitance between turns, and the primary and secondary would act like parallel tuned circuits at some frequency. Thus, the "untuned" amplifier would be, in effect, frequency selective to a degree depending largely on the nature of the coils used. Sometimes such an amplifier is used where high selectivity is not desired.

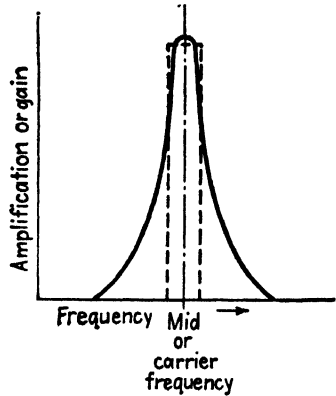



FIG. 173.—The dotted figure is the shape of the characteristic curve of an ideal radio-frequency voltage amplifier that is to pass a carrier and the two sidebands. The solid curve is the approximate shape of the characteristic curve that can be obtained in practice.

Where selectivity is desired and a definite frequency band is to be passed and other frequencies rejected, either the primary or the secondary is tuned by placing a variable condenser is parallel. By varying the condenser, the band of frequencies amplified and passed can be controlled. Of course it is possible to use a fixed condenser and to vary a small magnetic core often in the form of a "button" or small cylinder of powdered compressed magnetic material placed within the coil across which the condenser is connected. This would be a **tuned-radio-frequency amplifier,** and it follows the theory explained on page 271. In many radio-frequency amplifiers *both* the primary and the secondary are tuned. The selectivity of modern radio-receiving sets, that is, the ability to tune in the desired station and to reject all other signals, largely depends on these transformers.

Radio-frequency Voltage Amplifier with Tuned Secondary.—The basic circuit and the equivalent circuit for this amplifier are shown in Fig. 174. With an impressed radio-frequency voltage composed of the signals from many radio-transmitting stations, the amplified voltage μE_g also will be a complex voltage containing many signal frequency bands. (In an actual amplifier the grid

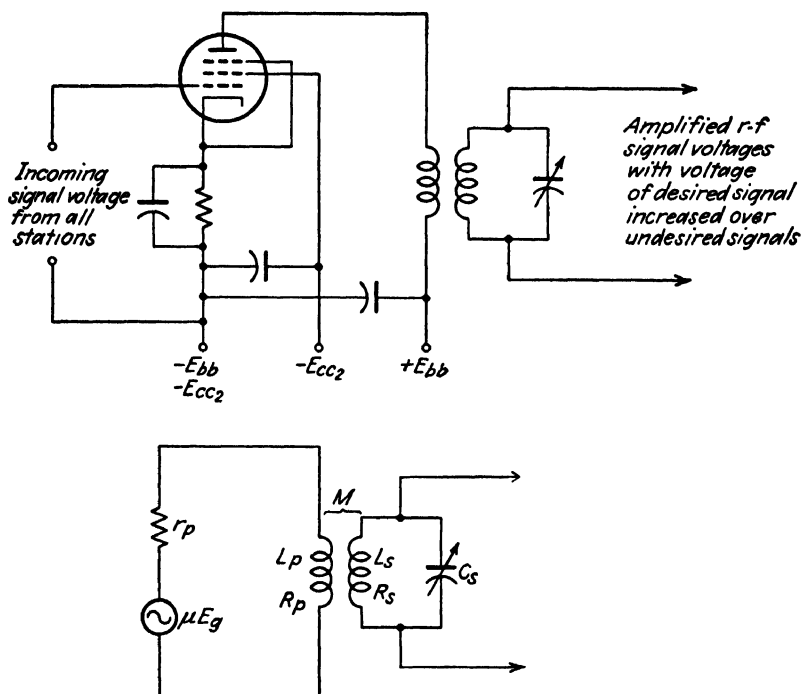



Fig. 174.—Actual and equivalent circuits for a radio-frequency voltage amplifier with transformer coupling and a tuned secondary. The letter symbol E_{cc2} represents the screen-grid supply voltage.

input circuit also would be tuned, but this was omitted for simplicity.)

The amplified signal voltage μE_g forces a complex current through the primary of the coil. The current components of various frequencies induce corresponding signal voltages *in series* in the secondary. These secondary voltages cause current components of corresponding frequencies to flow around the secondary. The $I_s X_{c_s}$ drop produced by each current component across the condenser C_s is the output voltage at that particular frequency.

To calculate the output voltage *at any frequency f*, first calculate the reflected impedance, using Eq. (33), page 102. For the equivalent circuit of Fig. 174 this will be

$$Z_{\text{reflect}} = \frac{(\omega M)^2}{Z_s} = \frac{(2\pi f M)^2}{R_s + j\,[2\pi f L_s - (1/2\pi f C_s)]}. \tag{93}$$

Next calculate the total primary impedance. This will be the sum of the primary and reflected impedances,

$$Z_{\text{pri. total}} = Z_p + Z_{\text{reflect}} = [(r_p + R_p) + j2\pi f L_p] + Z_{\text{reflect.}} \tag{94}$$

The current that flows through the primary will be

$$I_p = \frac{\mu E_g}{Z_{\text{pri total}}}, \tag{95}$$

and the magnitude of the voltage induced in the secondary will be

$$E_s = 2\pi f M I_p. \tag{96}$$

This voltage forces a current around the secondary that equals

$$I_s = \frac{E_s}{Z_s}. \tag{97}$$

and the voltage output is the drop across the condenser C_s, or

$$E_{\text{out}} = I_s X_{c_s}. \tag{98}$$

This gives the output *at any frequency f* if the magnitude of the voltage component of the frequency that is impressed on the grid is known. Of course Eq. (98) can be written as an equation combining all terms, but it is more meaningful to perform each step separately.

The voltage output at the frequency f_r for which the secondary is tuned also can be calculated as just explained. However, in this instance Z_s equals R_s approximately. Also with a pentode, r_p largely controls the primary current flow. With such assumptions the voltage gain *at resonance* is

$$A_v = g_m\, 2\pi f_r M Q, \tag{99}$$

where g_m is the mutual conductance of the tube in mhos, f_r is the frequency in cycles per second, M is the mutual inductance in henrys, and Q is the storage factor of the secondary (page 49).

Radio-frequency Voltage Amplifier with Tuned Primary and Secondary.—If a suitable condenser is connected across the primary of the transformer of Fig. 174, an amplifier with tuned pri-

mary and secondary results. This is a very important amplifier, and it is used in almost all modern radio-receiving sets. With a condenser connected across the primary, the equivalent circuit of Fig. 175 results.

If it is assumed that the amplified voltage μE_g is a complex voltage containing many frequency components, the output voltage *at any frequency f* can be found in the following manner: First calculate the impedance reflected by the secondary into the primary. This will be as given by Eq. (93). Next calculate the total impedance of the transformer primary including the effect of the secondary. This will be given by Eq. (94). It is then necessary to calculate [using Eq. (23), page 82] the equivalent series impedance of the condenser C_p in parallel with the *total* primary impedance, as just discussed. The total impedance into which the amplified voltage μE_g works will be this equivalent series impedance added to the plate resistance. The current flowing through r_p will be the voltage μE_g (for the frequency component f under consideration) divided by this total impedance, and the voltage across the transformer tuned primary will be μE_g minus the drop across the plate resistance. The current through the primary coil of the transformer will be the voltage across the transformer divided by the total primary impedance as given by Eq. (94). This current will induce [Eq. (25), page 86] a voltage in series in the secondary. This voltage divided by the total secondary impedance Z_s gives the secondary current, which when multiplied by the reactance of condenser C_s gives the output voltage *at any frequency f*. All operations must be vectorial in nature.

FIG. 175.—Equivalent circuit for a radio-frequency transformer-coupled amplifier with tuned primary and tuned secondary. This method of coupling is used extensively in radio-receiving sets.

The method just discussed is very laborious, and short cuts are possible, particularly at the frequency at which the primary and secondary are in resonance. Thus, with pentodes having very high plate resistances, the amplification or voltage gain A_v *at the resonant frequency* f_r becomes[1]

[1] The equations immediately following are from "Radio Engineers' Handbook," by F. E. Terman, McGraw-Hill Book Company, Inc.

$$A_v = \frac{g_m\, k\, 2\pi f_r \sqrt{L_1 L_2}}{k^2 + \dfrac{1}{Q_1 Q_2}}, \tag{100}$$

where g_m is the grid-plate transconductance, or mutual conductance, of the pentode expressed in mhos, k is the coefficient of coupling between the primary and secondary, f_r is the resonant frequency in cycles, L_1 and L_2 are the self-inductance in henrys of the primary and secondary, and Q_1 and Q_2 are the energy-storage factors of the primary and secondary (page 49).

As has been explained (page 308), a radio-frequency amplifier as used in the ordinary radio set must pass a band of frequencies about 20,000 cycles wide composed of the carrier and the two sidebands. An amplifier such as Fig. 175 follows the coupled-circuit theory of page 105. If the coupling is too loose, the band passed will be too narrow, and frequency distortion will result because the sidebands will be "clipped." If the coupling is too close, the band passed will be too wide, and the amplifier will not be sufficiently selective. The width of the band passed is related to the mean or carrier frequency f_c by the relation

$$\text{Width of band passed} = 1.2\, k f_c. \tag{101}$$

Substantially constant amplification over the band passed results when the Q_1 of the primary and Q_2 of the secondary, and the coefficient of coupling k have the following relation

$$\sqrt{Q_1 Q_2} = \frac{1.75}{k}. \tag{102}$$

Because an amplifier that has the equivalent circuit of Fig. 175 amplifies and passes a given band of frequencies and tends to reject all frequencies outside this band, it is often called a **band-pass amplifier.** From the discussion just given it is seen that these amplifiers follow coupled-circuit theory (page 105), and a review of this theory will be very helpful in predicting band-pass amplifier performance.

SUMMARY

Voltage amplifiers are used to increase a feeble signal voltage. Power amplifiers are used to drive some device, such as a loudspeaker or an antenna.

In the amplification of a complex signal, distortion may result. The three types of distortion are (*a*) frequency distortion in which the various frequency components are changed in their relative magnitudes, (*b*) nonlinear distortion

in which new frequencies are created, and (*c*) delay distortion in which the relative phase positions of the components are changed.

Voltage amplifiers usually are classified as audio-frequency and as radio-frequency amplifiers. A third classification is video amplifier, a type used in television. Audio-frequency amplifiers operate from about 50 to 10,000 cycles. Most radio-frequency amplifiers operate from about 100,000 cycles to hundreds of millions of cycles. The video amplifiers operate from about 30 to 3,000,000 cycles. A stage of amplification consists of a single tube and its input and output circuits. Two or more stages are coupled together with resistors and capacitors giving the resistance-coupled amplifier, or with transformers giving the transformer-coupled amplifier.

For voltage amplification the control grid of the tube is biased negatively, and the alternating signal voltage is connected in series with this bias The signal-voltage swing drives the control grid less negative or more negative, allowing more plate current or less plate current to flow through the load resistor. This alternating-current component causes an alternating voltage drop across the load resistor, and this is the amplified output voltage. The voltage gain of a single stage with resistance R_L in the plate circuit is

$$A_v = \frac{\mu R_L}{(r_p + R_L)},$$

and the phase shift between the input and output voltages is 180°.

In the design of resistance-coupled audio-frequency voltage amplifiers, the gain at intermediate frequencies (such as 1000 cycles) is calculated by the equation just given, where R_L is the equivalent resistance of the load resistor and the grid resistor in parallel. At low frequencies (such as 50 cycles) the gain and phase shift are given by Eq. (82). The low frequency at which the gain is 70 per cent of that at intermediate frequencies, and at which the phase shift is 45°, can be calculated by Eq. (83). At high audio frequencies (such as 10,000 cycles) the gain and phase shift are calculated by Eq. (84). The high frequency at which the gain is 70 per cent of that at intermediate frequencies, and at which the phase shift is 45°, can be calculated by Eq. (85).

For a resistance-coupled voltage amplifier the theoretical maximum gain occurs when the load resistance is very high compared to the plate resistance of the tube. For triodes the ratio of the load resistance R_L to the plate resistance r_p is from 3 to 5. For tetrodes the load and plate resistance often are of the same magnitude. For pentodes the load resistance usually is from one-half to one-tenth that of the plate resistance.

A typical resistance-coupled amplifier using triodes has moderate gain and amplifies over a wide band. A similar ampifier using pentodes has high gain but does not amplify over so wide a band.

Although the input impedance of a vacuum tube often is considered infinite, actually this is not true. A small amount of direct grid-current flows. With triodes having resistance loads, feedback of signal from the plate circuit to the grid circuit causes a high input capacitance. When a triode has an inductive load, power is fed back from the plate to grid and the tube tends to oscillate. Tetrodes and pentodes are not troubled with this, except at very high radio frequencies, because of the shielding of the extra grids between the plate and

control grid. At very high radio frequencies the transit time of the electrons in passing across the tube may cause the grid to draw power because of what is called "electron loading."

When several tubes operate from a common power supply, the output impedance of the power-supply filter is common to the tubes. If the amplified signal current component of the last tube is sufficient, the IZ drop that it causes across the common impedance may feed back signal voltage into the preceding tubes and may cause a type of very low-frequency oscillations called "motor-boating." This is very likely to occur in a three-stage amplifier. Decoupling resistors and capacitors minimize this tendency to motor-boat.

For pentodes the approximate voltage gain per stage can be found by the equation

$$A_v = g_m R_L,$$

but this equation should not be used with triodes. The equation previously given in this summary is based on Thévenin's theorem. The equation just given is based on Norton's theorem.

Compensated wide-band resistance-coupled voltage amplifiers are used as video amplifiers in television apparatus. By adding a parallel R-C network in the load circuit the low-frequency limit can be lowered. The high-frequency limit can be increased by adding a small coil in series with the load resistor.

With transformer-coupled audio-frequency voltage amplifiers using good interstage transformers the gain per stage is approximately

$$A_v = \mu N.$$

Radio-frequency voltage amplifiers may be used to amplify a single frequency only, but those in the conventional radio-receiving set must amplify a band of frequencies composed of the carrier and the two sidebands. For high quality amplitude-modulation radio programs the band width is about 20,000 cycles.

A radio-frequency voltage amplifier uses tetrodes or pentodes because these do not feed back signal and oscillate in the usual radio circuit. Tuned radio-frequency transformers are used between stages. Sometimes only the secondary of the transformer is tuned, but quite often both the primary and secondary are tuned.

For the radio-frequency amplifier with untuned primary and tuned secondary, the voltage gain per stage at resonance is

$$A_v = g_m 2\pi f_r MQ.$$

For the radio-frequency amplifier with tuned primary and tuned secondary the voltage gain per stage at resonance is given by Eq. (100).

REVIEW QUESTIONS

1. What is the fundamental difference between a voltage amplifier and a power amplifier?
2. Name the three types of distortion, and describe each type.
3. How are voltage amplifiers classified?
4. Why is the dynamic curve rather than the static curve used in Fig. 155?

5. Why is it possible to represent an amplifier tube as a generator of voltage μE_g and internal resistance r_p?

6. What is the phase relation between applied grid signal voltage and amplified output voltage for a tube with resistance load?

7. Refer to Fig. 158, and explain the functions of R_L, C, and R_g.

8. What does C'_g of Fig. 159 represent?

9. What element of a resistance-coupled amplifier circuit is very important in determining the low-frequency response? Why?

10. What is an important factor in determining the high-frequency response of a resistance-coupled amplifier? Why?

11. What should be the ratio of R_L to r_p for maximum theoretical voltage amplification? Why is this ratio not used? What ratios are used?

12. What are the functions of R_c and C_c of Fig. 164?

13. What are the functions of C_F and R_F of Fig. 164?

14. What are the important reasons that an extremely high grid resistor is not used?

15. Why are tetrodes and pentodes instead of triodes better suited for operation at radio frequencies?

16. Discuss the grid input impedances of tetrodes and pentodes as compared with triodes when used as amplifiers with resistance in the plate circuit.

17. When tetrodes are used in resistance-coupled amplifiers, the load resistance is about the same as the plate resistance, but when pentodes are used the load resistance is much less than the plate resistance. Why?

18. How does the gain per stage and the frequency response of a resistance-coupled amplifier using triodes compare with one using pentodes?

19. What is meant by a compensated amplifier? Why is compensation necessary and how is it achieved?

20. How is the gain per stage of a transformer-coupled audio-frequency amplifier calculated. Why may this method be used?

21. How are radio-frequency voltage amplifier stages coupled?

22. Discuss the nature of the signal that must be passed in the radio-frequency amplifier portion of a radio-receiving set for amplitude-modulated programs.

23. Why does a transformer-coupled radio-frequency amplifier *without* parallel tuning condensers on the primary and secondary exhibit the properties of a tuned amplifier?

24. Why must a radio-frequency voltage amplifier be highly selective?

25. Refer to Fig. 174*a*, and explain why condensers are connected from screen grid and plate to cathode. Would they be necessary in a single-stage amplifier? Why?

26. Refer to Fig. 174*b*, and explain how the amplified voltage μE_g produces an amplified output signal.

27. Why is the circuit of Fig. 175 superior to that of Fig. 174*b* for many radio-frequency purposes?

28. Refer to Fig. 175, and explain how the amplified voltage μE_g produces a signal voltage output.

29. On page 313 it is stated that the sidebands may be "clipped." What is meant by this statement, and what is the cause of clipping?

30. Why is an amplifier with tuned primary and secondary called a "band-pass" amplifier?

PROBLEMS

1. Repeat the calculations on page 273, using a plate load resistor of 300,000 ohms, and also calculate the gain for two stages.

2. Calculate the voltage gain per stage at the intermediate audio frequencies for a triode having a $\mu = 13.8$ and an $r_p = 12{,}000$ ohms when used with a load resistor $R_L = 50{,}000$ ohms, a coupling capacitor $C = 0.01$ microfarad, and a grid resistor $R_g = 500{,}000$ ohms. Also calculate the low frequency and high frequency at which the gain will be 70 per cent of the value at intermediate audio frequencies.

3. Refer to the illustrative problem starting on page 284, and calculate the required direct-current power dissipating capacity required for R_C, R_L, and R_F of Fig. 164. Specify what type of resistor should be used for each (see page 55).

4. If a triode has a $\mu = 9.3$, $r_p = 11{,}000$, $C_{gp} = 6$ micromicrofarads, and $C_{gk} = 3$ micromicrofarads, calculate the effective grid input capacitance when operated with an equivalent load resistance of 35,000 ohms.

5. Calculate the voltage gain per stage at intermediate frequencies for a pentode having a $\mu = 1500$ and an $r_p = 1{,}500{,}000$ ohms when used with a load resistor $R_L = 250{,}000$ ohms, a coupling capacitor $C = 0.005$ microfarad, and a grid resistor $R_g = 500{,}000$ ohms. Also calculate the low frequency and the high frequency at which the gain will be 70 per cent of the value at intermediate audio frequencies.

6. Refer to the illustrative problem starting on page 292, and calculate the required direct-current power dissipating capacity required for R_C, R_d, R_L, and R_F of Fig. 166. Specify what type of resistor should be used for each (see page 55).

7. Use the data of Prob. 2, and calculate the gain per stage, using Eqs. (80) and (87). Compare the results, and draw conclusions.

8. Calculate the over-all gain of the amplifier of Fig. 171 if the tubes used are type 6C5 having $\mu = 20$ and $r_p = 10{,}000$ ohms, and if the interstage transformers have a ratio of 1 to 2.7. Also calculate the value of R_c and the direct-current power-handling capacity it must have.

CHAPTER IX

POWER AMPLIFIERS

As was explained in the preceding chapter, there are two basic types of amplifiers, voltage amplifiers and power amplifiers. It was shown that the purpose of a voltage amplifier is to take the feeble voltage that is impressed on its input terminals and to amplify this voltage until it is sufficiently strong to accomplish some desired result, such as driving the grid of a power tube.

A power tube is fundamentally the same as a voltage-amplifying tube. A power tube is, however, usually larger in size, and is so constructed internally that the driving grid signal voltage, the applied plate voltage, and the resulting plate current may be much larger than with voltage-amplifying tubes. Whereas voltage-amplifying tubes will put out a small amount of power, power tubes will handle large amounts of power. For this reason, power tubes drive loudspeakers, and furnish the power input to radio-transmitting antennas.

The term "power amplifier" implies that power flows into the grid of the power tube, that this power is increased or amplified, and that this increased power then flows out the plate circuit to the connected load, such as a loudspeaker or an antenna. Sometimes this is true, but often the grid-input power is negligible; it depends on the type of power amplifier. Be that as it may, the name is in general use, and will be used in this text to describe circuits whose main purpose is to deliver signal power to connected loads.

Power Transfer and Impedance Matching.—These subjects were considered in detail in Chap. IV, but are so very important in power-amplifier design that they will be reviewed.

In the preceding chapter it was shown that a voltage-amplifying tube could be considered as a generator with an open-circuit or generated voltage of μE_g and an internal resistance of r_p. This simple equivalent circuit also holds for many power amplifiers, in fact, it can be said to hold for all power amplifiers, although as will be seen, certain power amplifiers are more conveniently studied by other methods.

Thus, when an alternating signal voltage of E_g volts is impressed on the control grid of a power tube, the power tube may be considered as a generator of voltage μE_g and internal resistance r_p. If a resistance load R_L is connected to such a tube, the current that flows is $I_p = \mu E_g/(r_p + R_L)$ in accordance with Eq. (78). The power delivered to the load will be

$$P = I_p^2 R_L = \left(\frac{\mu E_g}{r_p + R_L}\right)^2 R_L. \tag{103}$$

If the discussion on page 99 is consulted, it follows that *maximum power* will be delivered by a power tube to a resistance load when the load resistance R_L equals the internal resistance r_p. A word of warning must be given at this point. In power amplification, signal distortion must be considered. In *some types* of power amplifiers, the nonlinear distortion (creation of harmonics) is excessive when $R_L = r_p$. It so works out that in *these particular* amplifiers $R_L = 2r_p$ gives approximately *maximum undistorted power output.* The word "undistorted" means that the distortion is less than about 5 per cent, an amount that for many purposes is not bothersome. The fact remains, however, that many power amplifiers give maximum power output when the load resistance equals the internal resistance of the tube, but excessive distortion may result.

Of course there is no object in having a power tube deliver its output to a resistor, except perhaps for test or other special purposes. The purpose of a power amplifier is to deliver signal power to some device, such as a loudspeaker or an antenna.

Usually, the input impedance characteristics of the devices to be driven are such that these should *not* be connected *directly* to the amplifier if maximum power transfer is desired. It must be realized that loudspeakers, transmitting antennas, and other devices to be driven are constructed as they are for certain reasons. Thus the dynamic loudspeaker has a certain form for mechanical as well as electrical reasons. Likewise, a radio-transmitting antenna may be made in a particular way, not because it gives a certain desired input impedance, but because it gives a certain radiation pattern, or because for economic reasons it can only be so many feet high.

In general, therefore, some impedance-matching network (page 109) must be inserted between the power tube and the power-consuming load, such as a loudspeaker or transmitting antenna.

For audio-frequency power amplifiers, this impedance-matching network usually is an audio-frequency power-output transformer with a core of magnetic material and a turns ratio different from unity so that the load impedance can be matched to the tube. For radio-frequency power amplifiers, the impedance-matching network may be in the form of an inductively coupled air-core transformer (page 101), or perhaps more commonly some impedance-transforming network of inductors and capacitors. In radio these often are referred to as **tuning circuits,** but impedance transforming and matching are the real purposes of these circuits.

In the following pages the power tubes often will be shown, and considered, as working into a resistance load R_L. Remember, therefore, that this load resistance R_L is a fictitious, or imaginary, resistance that exists not in fact but in effect: It is the alternating-current resistance that is transformed or coupled into the plate circuit by the impedance-matching network.

Types of Amplifier Operation.—The tubes used in some amplifiers are operated on the straight portion of their dynamic characteristic curves. This is not true for all amplifiers, however. The mode of operation of the tube determines one important means of classification.

Class A Amplifier.—When a tube is operated in class A, the grid bias and the alternating grid signal voltage are such that plate current flows at all times. The amplifier of Fig. 155, and all other amplifiers thus far considered, were operated in class A. Such amplifiers can be operated so as to produce but little distortion, and since in voltage amplification this is very important, voltage amplifiers are operated in class A.

In class A operation (Fig. 176) it is common practice to arrange the direct grid bias E_c so that the point of operation P is on the approximately straight part of the dynamic characteristic curve. Also, it is common practice that the peak of the alternating grid-signal voltage, or **grid swing,** as this signal voltage is called, does not exceed the grid-bias voltage E_c. With such operation the grid never is driven positive. But such operation is not necessary for classification as a class A amplifier. According to the definition, all that is required is that the instantaneous total plate current i_b never falls to zero.

Thus the grid may be driven positive, and if this happens, grid current flows. The designation class A1 indicates that no grid

current flows, and class A2 indicates that grid current flows. Most class A amplifiers operate as class A1, with negligible grid current flowing (see, however, page 374). When the designation class A is used, it often implies class A1 operation.

Class B Amplifier.—When a tube is operated in class B, the grid bias is equal approximately to the cutoff value, the plate current

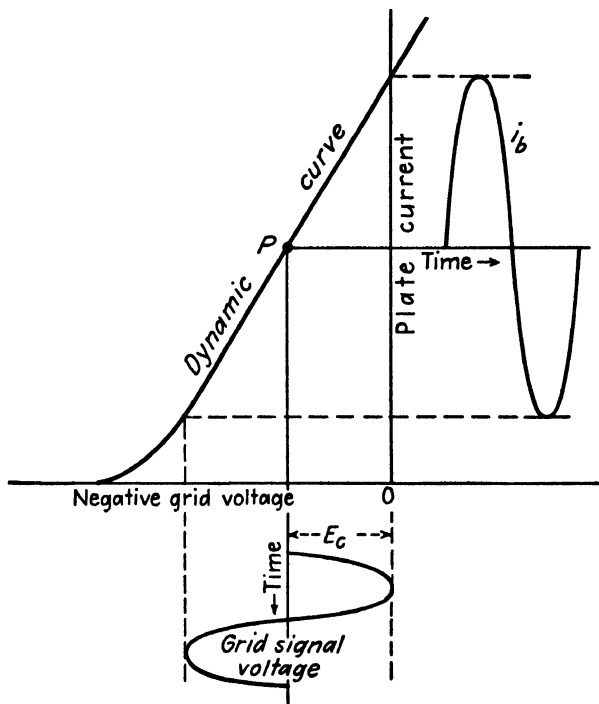



FIG. 176.—Class A1 operation. The point of operation P is near the center of the approximately straight portion of the dynamic curve. This point P is determined by the magnitude of the negative grid-bias voltage E_c, and by other conditions of operation. For class A2 operation, the grid signal voltage would be increased until the grid is driven positive and appreciable grid current flows.

is approximately zero with no applied alternating grid signal voltage, and plate current flows for approximately one-half of each cycle when an alternating grid signal voltage is applied. Such operation is indicated in Fig. 177.

With such operation, plate current flows only for the positive half cycles of the applied grid signal voltage. If the grid is not driven positive and negligible grid current flows, the designation is class B1; if the grid is driven positive and grid current flows, the

designation is class B2. The designation class B often implies class B2 operation.

Class C Amplifier.—When a tube is operated in class C, the grid bias is appreciably greater than the cutoff value, the plate current is zero with no applied alternating grid signal voltage, and plate current flows for appreciably less than one-half cycle when a grid signal voltage is applied. Such operation is indicated in Fig. 178,

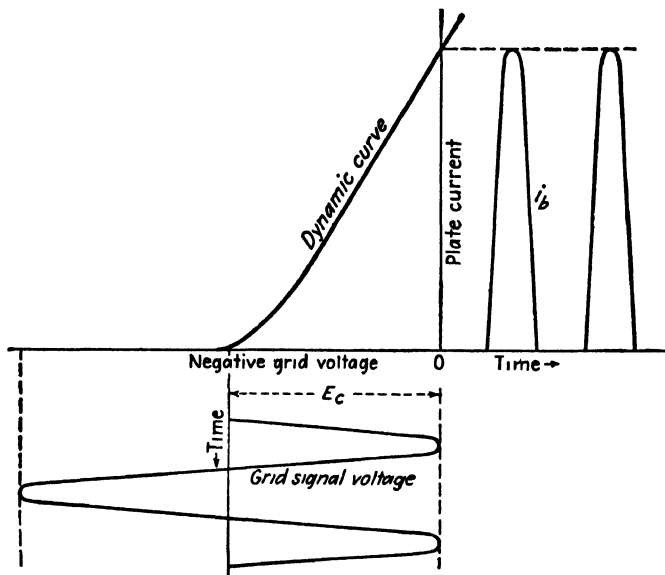



FIG. 177.—Class B1 operation. The tube is biased to cutoff by the negative grid-bias voltage E_c. For class B2 operation, the grid-driving voltage would be increased until the grid is driven positive and appreciable grid current flows.

where the grid is biased by the direct voltage E_c to about twice the cutoff value.

For class A operation, plate current always flows, and in terms of the applied grid signal voltage (assumed to be a sine wave), plate current flows for the entire cycle or 360°. For class B operation, plate current flows for approximately one-half cycle, or for 180°. For class C operation, plate current flows for less than 180°, as indicated by θ_b of Fig. 178. This is because plate current flows only when the alternating signal voltage drives the grid less negative than the cutoff value. The **angle of flow** θ_b accordingly is less than 180°. In Fig. 178 the grid is driven positive (although of course this is not necessary for class C operation), and when the

grid is driven positive, grid current flows. It will be noted that the angle of flow θ_c for the grid current is less than θ_b for the plate current. The designations class C1 and class C2 also are used as previously explained.

Operation sometimes is designated as class AB when it is between the class A and class B operation.

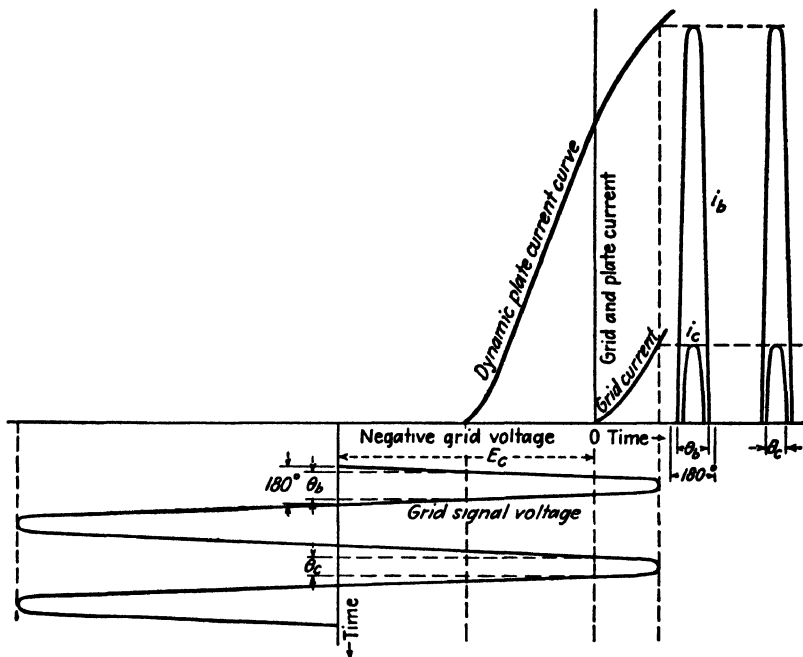



FIG. 178.—Class C2 operation (usually called class C operation). The tube is biased far beyond cutoff by the negative grid-bias voltage E_c. As shown here, the bias is about twice cutoff; often it is much greater. This is called C2 operation because grid current flows.

Additional Classifications of Power Amplifiers.—Based on the frequency of the signals amplified, power amplifiers may be classified as **audio-frequency power amplifiers** or as **radio-frequency power amplifiers.**

With minor exceptions, power amplifiers employ triodes or pentodes. The beam-power tube (page 197) is here classed as a pentode. These tubes have low amplification factors and low plate resistances as compared with corresponding voltage-amplifying tubes. Also, power tubes usually are larger, more rugged, and operate with higher electrode voltages and plate currents than

voltage-amplifying tubes. A possible classification is **triode power amplifiers** or **pentode power amplifiers.**

A power amplifier often consists of a single power-output tube, and is sometimes called a **single-ended amplifier,** a term none too descriptive. Often, a power amplifier has two tubes and is operated as a **push-pull amplifier,** a term that will be explained on page 333. In some instances two or more power-output tubes are operated in parallel, and sometimes in push-pull parallel.

In this chapter audio-frequency (a-f) power amplifiers will be discussed first, and radio-frequency (r-f) power amplifiers will then be considered.

Audio-frequency Single-triode Class A Power Amplifier.—A typical arrangement of a power amplifier following a voltage amplifier is shown in Fig. 171, page 303. The amplified signal voltage is impressed on the grid of the power tube through an audio-frequency input transformer, and the amplified power output is coupled to the load through a power-output transformer (page 303). Two ways are commonly used to explain the operation of this amplifier, (*a*) the dynamic-curve method, based on constant plate-voltage static curves, and (*b*) the load-line method, based on constant grid-voltage static curves. These methods will now be considered.

The Dynamic Curve.—If there is no resistance in the plate circuit, then all the voltage impressed between the cathode and plate actually appears between the cathode and plate, and *the plate voltage remains constant* when a signal is impressed on the grid. But, as has been explained elsewhere, an amplifier tube without a load in the plate circuit cannot amplify, and, hence, tubes have either a resistance connected into the plate circuit or a resistance coupled into the plate circuit by the power-output transformer or coupling network. Then when a signal voltage is impressed on the grid, the plate current varies in accordance (page 270), and the plate voltage, as measured between plate and cathode, *does not remain constant,* but varies also with the signal.

It follows, therefore, that static curves, taken with *no load* in the plate circuit cannot be used *directly* to predict amplifier operation. The so-called **dynamic curve,** taken with resistance in the plate circuit, must be used. But this would mean that tube data, as included in tube manuals and elsewhere, would have to include dynamic-curve data for *each possible load,* which is not practicable. Instead the data for the static curves are given, and from these

the dynamic curve, for operation with a given load resistance, can be computed. (As a matter of fact, tube data usually are given in the form of constant grid-voltage curves, but these data readily can be plotted as constant plate-voltage curves if desired.)

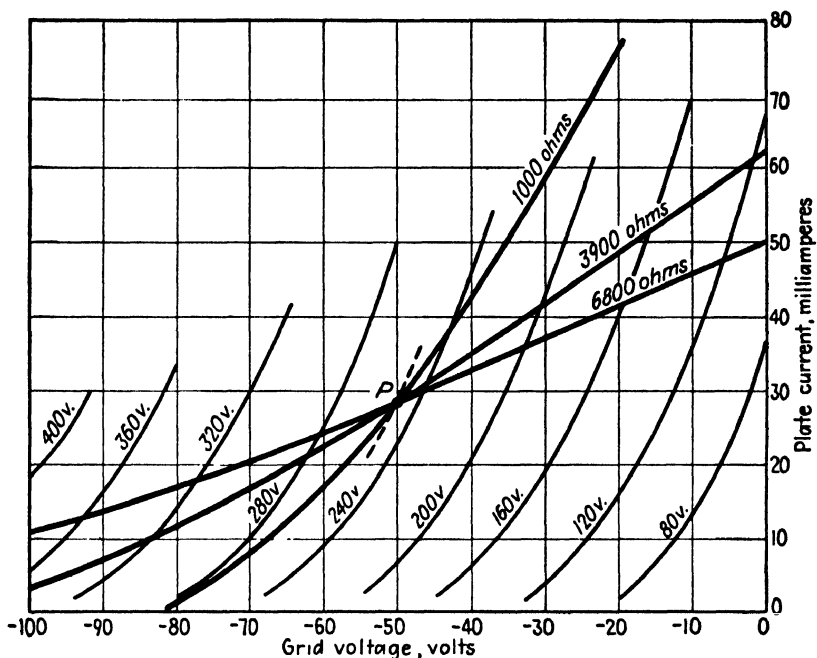



FIG. 179.—Static and dynamic curves for a small power-output triode. If the tube is operated without resistance in the plate circuit, then the plate current varies in accordance with the static curves. With resistance in the plate circuit (as usually operated), the plate current varies in accordance with the dynamic curves. As an illustration of how to compute a dynamic curve, the 3900-ohm curve will cross the 120-volt static curve at a current value $I = [28 + (250-120)/3900] = 61$ milliamperes. The 28 milliamperes is the current at the point of operation P. The 3900-ohm curve will cross the 320-volt static curve at the current value $I = [28 - (320-250)/3900] = 10$ milliamperes. The dynamic curve may be obtained experimentally by taking the characteristics of the tube with resistance inserted in the plate circuit. At the operating point P, the plate voltage is $+250$ volts and the grid voltage is -50 volts. Other electrode potentials and points of operation may be used.

Constant plate-voltage curves for a small power-output triode are shown in Fig. 179. How dynamic curves are determined from static curves is explained beneath this figure. When this tube is operated with 250 volts on the plate and a grid bias of -50 volts, the amplification factor is 3.5 and the plate resistance is 1610 ohms. For operation as a single class A audio-frequency power amplifier,

the plate-load resistance is 3900 ohms for maximum undistorted power output (page 331), and the peak value of the applied alternating grid signal voltage (or grid swing) is 50 volts.

Because a tube with resistance in the plate circuit operates in accordance with the dynamic curve, Fig. 180 has been included.

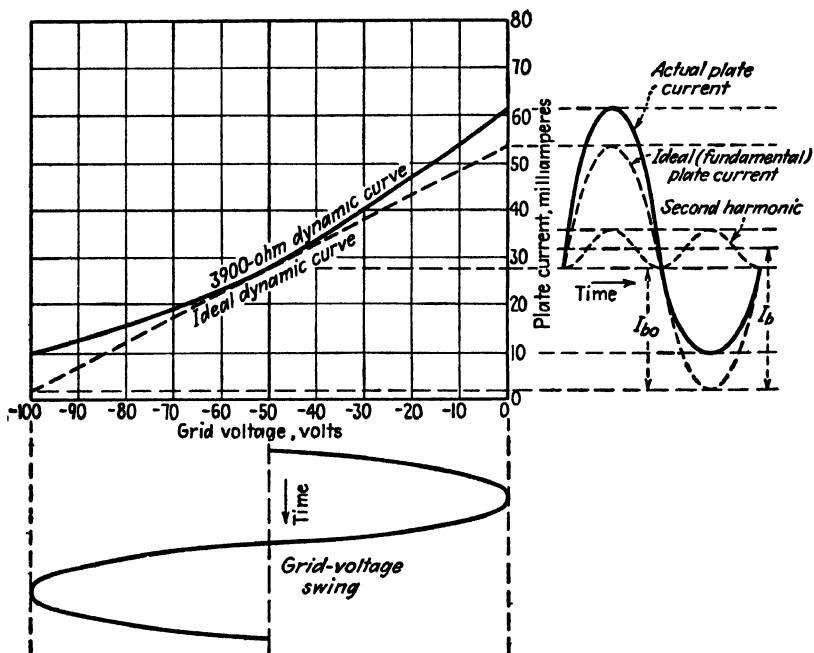



FIG. 180.—The curvature of the dynamic curve causes distortion, largely consisting of the creation of a second harmonic. If the dynamic curve were straight, then the plate current would be determined by the ideal dynamic curve, and the plate current would be undistorted, containing only the fundamental. In actual operation, the "fundamental" would be the signal wave.

As is indicated, if operation were along the "ideal dynamic curve," no distortion would result; but since actual operation is along the dynamic curve, nonlinear distortion results. This distortion largely consists of a second harmonic, although other frequencies also are created. A study of the ideal sinusoidal plate current and the actual plate current will indicate that a rather large second-harmonic current must exist in the actual output wave. It will be noted that when the signal is applied the direct-current component I_{bo} changes to a larger direct current I_b. Thus when signal is applied, the direct plate current as read by a milliammeter in-

creases in amount proportional to the magnitude of the second-harmonic distortion. This is a simple way to detect the presence of excessive nonlinear distortion if an oscilloscope or a wave analyzer is not available.

The dynamic curve is very convenient for showing the causes of nonlinear distortion. These are illustrated in Fig. 181 with accompanying discussions. It is of interest to note that in *voltage* amplifiers, frequency distortion is of most concern. This is because the signal voltage being handled by a tube usually is low and nonlinear distortion is accordingly negligible. On the other hand, *power* tubes usually are worked to the limit. For economic reasons a large power tube would not be purchased and installed, and then worked to only a fraction of its capacity. The practical solution is to use as small a power tube as reasonable, and work it to full capacity. Operation beyond this limit may cause nonlinear distortion, as indicated in Fig. 181.

The distortion considered in Fig. 181*b* does not, in a sense, occur in the tube, but rather in the source of signal voltage driving the tube. If the grid is driven positive on the positive half cycle of the applied signal, the grid draws current and this current flows through the source of signal voltage. If this source has internal impedance, as it usually does, then an internal drop of voltage occurs for the part of the cycle that the grid is positive. The internal drop subtracts from the total signal voltage, and thus the actual voltage that reaches the grid on the peak of the positive half cycle is "rounded off" and distorted.

This is one of the important reasons that the grids of tubes often are so biased that they never are driven positive. If they are driven positive, then the source of driving voltage must have power-output capacity (which is equivalent to saying low internal impedance) so that the terminal voltage of the source does not fall when grid current flows. This is very important when operation is in class A2, class B2, or class C2, and grid current flows.

The Load Line.—Although dynamic curves are very useful for purposes of explanation, load lines are more useful for quantitative study and amplifier design. Also, a load line is easier to determine because it need not be drawn point by point as must be done for the dynamic curve. This fact is illustrated by Fig. 182.

If the tube has *no resistance* in its plate circuit and has a bias of −50 volts and a plate voltage of 250 volts, and if a grid signal

voltage swings the grid from zero to −100 volts, then the plate current for the positive half cycle would rise to a high value, as determined by the point where the vertical line intersects the

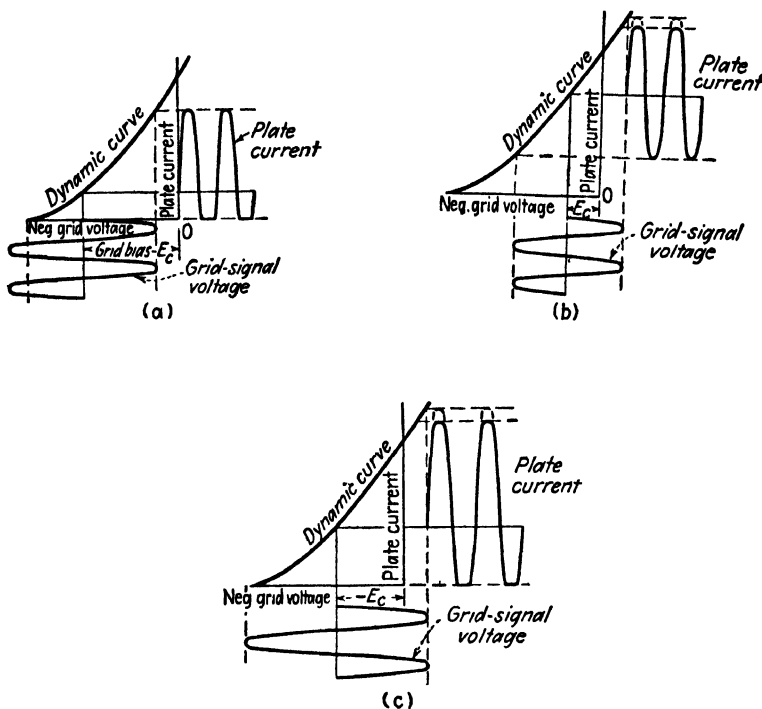



FIG. 181.—Showing how distortion is caused by improper operation of vacuum tubes. In (*a*) the grid bias $-E_c$ is too great, and the grid is driven beyond cutoff, thus distorting the negative half cycle. In (*b*) the negative grid bias E_c is too small and the grid will draw current from the source of grid signal voltage. This will cause a voltage drop in the internal impedance of the source. Then, instead of the actual signal voltage appearing on the grid, the voltage that appears will be the signal voltage minus the drop in the source, and hence the grid signal voltage will not be the true undistorted signal voltage. This action occurs on the positive half cycle when grid current flows, and the positive half cycle is distorted as shown. In (*c*) is shown the effect of too large a grid signal voltage. By the combined action of (*a*) and (*b*) both half cycles are distorted.

$E_c = 0$ curve, and for the negative half cycle would fall to zero, because the tube would be driven beyond cutoff. If, on the other hand, a tube had a *very large* resistance in its plate circuit, the current changes would be very small because the generated voltage μE_g could force but little alternating current through the tube.

Thus operation with *no load resistance* would follow the vertical line extending upward at a plate voltage of 250 volts. A projection to the left of the points of intersection with the grid-voltage curves would give the magnitude of the maximum and minimum plate current. These projections have not been shown for the sake of simplicity. Operation with *large load resistance* would be

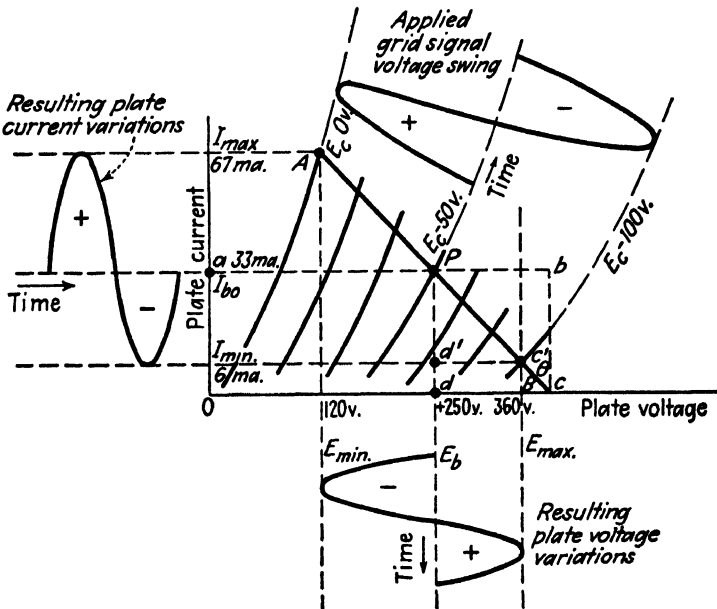



FIG. 182.—Diagram illustrating how the load line *A-B* is used to determine the performance of a triode power-amplifier tube. These are constant grid-voltage curves. Note that at high values of plate voltage the plate current is kept at low values to prevent overheating the plate. A few electrons at high velocity (caused by a high plate voltage) can produce as much heating at the plate as many electrons at low velocity resulting from low plate voltage.

determined by a line that was almost horizontal because, as explained in the preceding paragraph, the current changes would be small. These have been omitted for the sake of simplicity.

A vertical line represents operation with zero resistance, and a line that is almost horizontal represents operation with large resistance. It becomes apparent, therefore, that the magnitude of the resistance determines the *slope* of the load line, and that this line determines the magnitudes of the maximum and minimum

values of plate current produced by a given grid-voltage signal swing. A typical value of load resistance is given by the line A-B of Fig. 182. If the slope is defined as the tangent of the angle θ that the line makes with the *vertical*, then tangent θ equals the side opposite divided by the side adjacent, or the length $E_{max} - E_{min}$ divided by $I_{max} - I_{min}$, which for Fig. 182 is numerically $(360 - 120)/(67 - 6) = 3900$ ohms (approximately). This unit is used because volts/amperes = ohms. This value of 3900 ohms is the value that was used for the dynamic curve which gave maximum undistorted power output. Thus the operations as explained by Figs. 179 and 182 are but different aspects of the same thing.

As previously explained, Fig. 182 is easier to develop, because to determine operation with a given load resistance it is necessary only to draw through the point of operation a line that has a slope equal to the load resistance to be studied. Then, when the grid swing is plotted, the resulting alternating plate current and alternating plate voltages can be determined. Also, *area* on a diagram such as Fig. 182 equals the product of volts and amperes, which is watts, hence, area under the curves represents power. For instance, rectangle *Oabc* represents the total direct-current power that would have to be delivered to the tube if it actually had a resistor in the plate circuit. Of this area, rectangle *OaPd* represents the power delivered to the internal plate circuit of the tube, and area *dPbc* represents the direct-current power lost in the load resistor. If, however, a transformer with little primary direct-current resistance were used to couple the alternating-current load into the plate circuit, then area *dPbc* largely would disappear. The alternating-current power delivered by the tube to this load is represented by the triangle *d′Pc′*. The **plate-circuit efficiency** would be the ratio of the area of triangle *d′Pc′* to rectangle *OaPd* (that is, the ratio of the signal power output to the direct power input). It is apparent that a tube in class A has low efficiency. The entire analysis just made assumes that the distortion is negligible.

Method of Design.—The problem usually is this: the power that must be delivered to some load is known. A certain power-output tube must be selected for this purpose, and the operating potentials must be determined. The correct value of load resistance into which the tube works must be found, so that maximum undistorted power output will be obtained from the tube. Knowing

all these interrelated factors, the turns ratio of the audio-frequency output transformer can be found. The method of design of an audio-frequency class A power amplifier using a single triode now will be considered.

Illustrative Problem.—An audio-frequency power amplifier is to deliver about 1.5 watts of essentially undistorted signal to a loudspeaker that has a voice coil of 8 ohms, essentially pure resistance. A tube is to be selected for this purpose, and all operating conditions are to be determined.

Solution.—Step 1. In a tube manual it is found that the type 45 triode will deliver an undistorted power output of 1.6 watts, when operated as a single class A1 amplifier with +250 volts on the plate and −50 volts on the grid, and that when so operated it has an amplification factor of μ = 3.5 and a plate resistance r_p = 1610, and the plate current is 34 milliamperes. In making amplifier calculations, it is possible to start at the beginning and to determine all data by a cut-and-try process. Usually much of the data are taken from the tube manual. These data include the load resistance that should be in the plate circuit and the value of the biasing resistor. The data given in the tube manual will be verified and the method of making the calculations will be explained.[1]

Step 2. From an examination of the static curves of a type 45 power-output triode and from a consideration of the power supply needed, it is concluded that a reasonable operating plate voltage is 250 volts. For a *single triode power amplifier*, the proper zero signal bias is

$$E_c = \frac{0.68\,E_b}{\mu} = 0.68 \times 250/3.5 = 48.7 \text{ volts.}$$

When alternating current is used to heat the filament, it should be increased by about half the voltage rating of the filament. The type 45 uses 2.5 volts on the filament, so that a grid bias of −50 volts will be used. With 250 volts on the plate and −50 volts on the grid, the static curves give a plate current of about 30 milliamperes. But since the rating given in the manual is 34 milliamperes, this value will be used. (When a signal is placed on a power triode the plate current rises slightly.) If 34 milliamperes are used as the plate current, the cathode self-biasing resistor should have a resistance of $R_c = E_c/I_b$ = 50/0.034 = 1470 ohms, which checks the tube manual.

Step 3. The load resistance to give 1.6 watts undistorted power output will be checked. (Of course, only 1.5 watts is needed, but if a tube will supply 1.6 watts, it certainly will provide the desired 1.5 watts; after the amplifier is built, the actual power output depends on the signal voltage impressed on the grid.) The tube manual gives 3900 ohms as the plate load for 1.6 watts output. If this were not known, it would have to

[1] The explanations will follow those for a type 45 triode, as given in the RCA Receiving Tube Manual RC-14. For simplicity, certain refinements have been omitted. The method given in this illustrative problem is of sufficient accuracy for most purposes.

be guessed at, and the cut-and-try method used until it was determined. As a first guess, probably 3200 ohms would be used, because of the rule, stated on page 319, that for maximum undistorted power output the load resistance should be about twice the plate resistance, and the plate resistance as given in the manual is 1610 ohms (Step 1). The power output and distortion with 3900 ohms in the plate circuit will now be checked. The load line (Fig. 182) having a slope (page 329) corresponding to 3900 ohms is drawn through the operating point of $E_b = 250$ volts and $E_c = -50$ volts. Now when the grid signal swings the grid to zero, the plate current will rise to a maximum value and will be about $I_{max} = 67$ milliamperes. When the grid-signal voltage swings the grid to twice the bias value, or -100 volts, the plate current falls to a minimum, or about $I_{min} = 6$ milliamperes. It will be assumed that the average of the plate current is $I_b = 34$ milliamperes as given in the tube manual. (The curve indicates that it would be less, but the direct-current value rises when signal is applied, as indicated in Fig. 180 and the accompanying discussion.) Also, when the grid is driven to zero, the plate voltage falls to 120 volts, and when the grid voltage is driven to -100 volts, the plate voltage rises to 360 volts. Since power is effective value of plate current times effective value of plate voltage, when E is in volts and I is in amperes, the power delivered by the tube to the 3900-ohm load resistance is

$$P = \frac{(I_{max} - I_{min})\ (E_{max} - E_{min})}{8}$$

$$= \frac{(0.067 - 0.006)\ (360 - 120)}{8} = 1.83 \text{ watts},$$

a value that agrees reasonably well with that given in the manual.

Step 4. The percentage of distortion with a load resistance of 3900 ohms and the tube delivering full output will now be checked. The distortion is assumed to be caused entirely by the second harmonic, and it would be the magnitude of the second harmonic divided by the magnitude of the fundamental and expressed as a percentage. For the tube under consideration this is

$$\text{Percentage of second harmonic} = \frac{[(I_{max} + I_{min}/2)] - I_b}{I_{max} - I_{min}} \times 100$$

$$= \frac{[(0.067 + 0.006)/2] - 34}{0.067 - 0.006} \times 100 = 4.1 \text{ per cent.}$$

Thus it is seen that with 3900 ohms in the plate circuit, and with the tube driven to maximum output for class A1 operation with the peak of the grid signal voltage equal to the grid bias, the nonlinear distortion is less than 5 per cent, the usual allowable limit. If it had not been known that the load resistance should be 3900 ohms, and if some other resistance such as 3200 ohms (twice the plate resistance) had been tried, then the power output might not be correct, and the distortion might be excessive. In such instances, a new load resistance would be assumed, a new load line established, and a new check would be made until the desired value was found by this cut-and-try procedure.

Step 5. With a loudspeaker load resistance of 8 ohms to be driven, and the tube to work into 3900 ohms, the proper transformer ratio is

$$N = \sqrt{\frac{Z_1}{Z_2}} = \sqrt{\frac{3900}{8}} = 22.1.$$

Exactly this ratio need not be used, but the transformer must be designed to work from about 4000 to 8 ohms, must have sufficient power-handling capacity, and must be able to carry the direct plate current of 34 milliamperes without saturating the core. The derivation of the power-output equation used in Step 3 is rather simple. The distortion equation used in Step 4 is somewhat more difficult to derive. The derivation is based on Fig. 182, and can be found in most books on electronics.

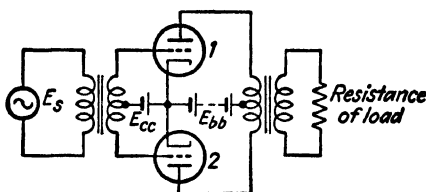



FIG. 183.—Basic circuit for two power-output triodes in a push-pull power-amplifying circuit.

Triodes in Parallel.—Sometimes triodes are operated in parallel to obtain greater power output. If two tubes are operated in parallel the load resistance into which the parallel plates should be connected is about one-half that for a single tube as previously determined. The power output of the two tubes is about twice that of a single tube.

Audio-frequency Class A Push-pull Power Amplifier.—If two triodes are operated in push-pull as in Fig. 183 instead of in parallel as discussed in the preceding paragraph, then the power output for comparable conditions of operation can be greater than twice that of a single tube. In push-pull, the impressed signal voltage drives one grid more positive and the other grid less positive on one half cycle, and for the next half cycle the action is opposite.

There are several advantages to push-pull operation: (*a*) The direct plate-current component of each tube flows in opposite directions through identical halves of the primary of the output transformer; and if these direct-current components are equal in magnitude, which usually is true, their magnetizing effects cancel, and they do not produce a constant magnetic flux in the core. As a result, the magnetic core may be smaller and the transformer will be lighter and less expensive than for two tubes in parallel

where the direct currents would combine and tend to saturate the core. (*b*) The second-harmonic-distortion components, generated

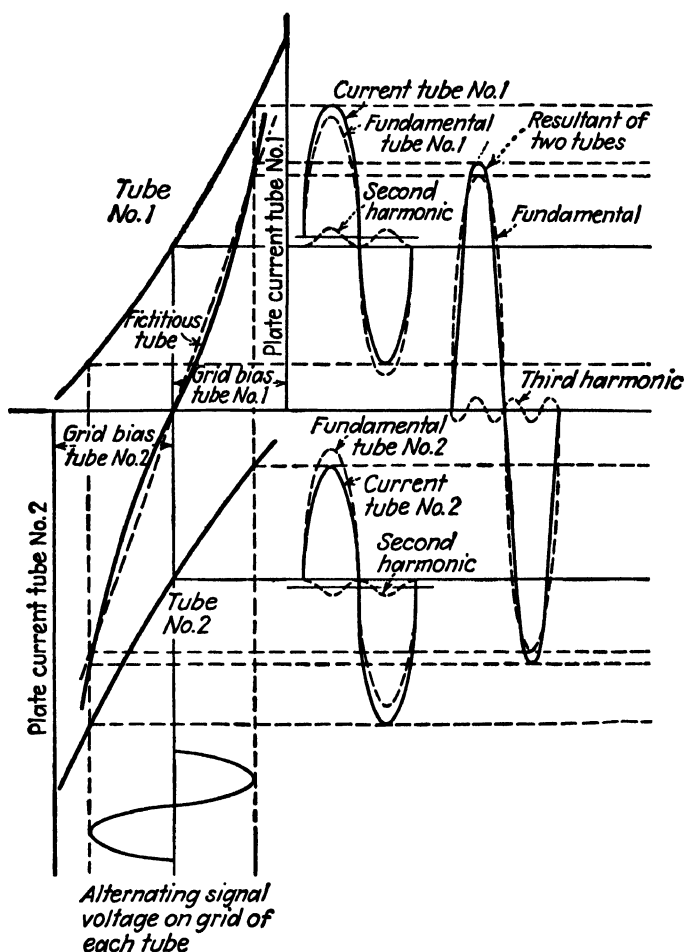



FIG. 184.—Diagram for studying two triodes in push-pull. Because of the curvature of the dynamic curves, second harmonics are created. These are out of phase and cancel in the transformer primary. The fundamental components act in phase, and their magnetic effects add in the output transformer. The effect is as if the resultant amplified signal current came from one fictitious tube having the characteristics shown. In this figure it can be thought that the fundamental current component is the result of impressing the signal voltage on the fictitious tube. Because the dynamic characteristic of the fictitious tube is not a straight line, third harmonics are created, and these do *not* cancel.

in the tubes, largely cancel, and thus they need not be considered in the design of the amplifier. In fact, all even harmonics cancel,

but the second harmonics usually are the only ones that are bothersome. Because of this, the selected plate load resistor *does not* follow the rule of being about twice the plate resistance, and the circuit may be designed so that a larger amount of power is drawn from the tubes. This is explained by Fig. 184. Another advantage is that hum from power supply, which is introduced at E_{bb} or E_{cc} of Fig. 183, is canceled. Hum that enters through the input transformer would be amplified just as any other signal.

FIG. 185.—Curves for determining power output and correct load resistance for push-pull triodes. The plate voltage is E_b = 250 volts. (*From RCA Receiving Tube Manual RC-14.*)

To determine the maximum power output of two triodes in push-pull, Fig. 185 is used.[1] By using a set of static curves, a vertical line is erected at 0.6 E_b, the direct plate operating voltage. The intersection of this with the zero grid-voltage curve (it is assumed that the grid of the tube is being driven by the signal voltage to zero and to twice the bias value, and it is assumed that operation is in class A1) gives the maximum current that flows. For the triodes of Fig. 185, which are the same tubes as in Fig. 182, the power output is

$$P = \frac{I_{\max} E_b}{5} = \frac{0.096 \times 250}{5} = 4.8 \text{ watts.}$$

This is considerably more than twice the value obtainable from one tube, as discussed in the preceding section. It is assumed that this power is without distortion, but of course some nonlinear distortion would exist, as Fig. 184 indicates. However, the quality of the output of a push-pull amplifier using triodes and well-designed transformers is excellent.

To determine the plate-load resistance that should be used, a line is drawn, as indicated, from the intersection of the vertical line and the zero grid-voltage curve to the direct plate-operating voltage

[1] This discussion is based on the method given in the RCA Receiving Tube Manual RC-14. For a theoretical discussion, consult B. J. Thompson, Graphical Determination of Performance of Push-pull Audio Amplifiers, *Proceedings of the Institute of Radio Engineers*, April, 1933.

E_b, as shown in Fig. 185. The slope of this line is the proper load resistance to use with a fictitious tube that would produce the resultant fundamental wave of Fig. 184. But, because the two push-pull tubes each work into *one half* of the primary, the proper *plate-to-plate* load resistance is four times that for the fictitious tube as shown by the slope of the line in Fig. 185. Thus the proper plate-to-plate load resistance is

$$\text{Plate-to-plate load resistance} = \frac{E_b - 0.6E_b}{I_{\max}} \times 4 = \frac{250 - 150}{96} \times 4 = 4160 \text{ ohms.}$$

If the push-pull amplifier is to drive an 8-ohm loudspeaker, the transformer turns ratio should be $N = \sqrt{Z_1/Z_2} = \sqrt{4160/8} = 22.8$.

Nonlinear distortion has been mentioned several times in connection with power transformers but nothing has been said about frequency distortion. This is because power tubes usually are worked to the limit, and nonlinear distortion is likely to occur (Fig. 181). Power-output transformers usually have good frequency response. One reason is that the number of primary turns need not be too high because these transformers work with power tubes having low plate resistances; another reason is that they usually work into low-impedance loads, and, hence, need but few secondary turns. This means that the distributed capacitances within the transformer are not as bothersome as they would be if the output transformers worked between high-impedance circuits.

Audio-frequency Single Pentode Class A Power Amplifiers.— As has been previously mentioned, tetrodes seldom are used for power amplifiers. Pentodes, and beam-power tubes that have essentially the same characteristics, are extensively used for power amplifiers. These power pentodes differ from voltage-amplifying pentodes, not in their general arrangement, but in the details that permit the passage of larger currents at higher voltages so that considerable signal power can be delivered to the connected load. Thus the coefficients of a power pentode are not so high as for a voltage-amplifying pentode; yet they are much higher than for a power triode of comparable size. A typical small power-output pentode, the 6F6, at 250 volts on the plate and -16.5 volts on the grid, has a plate current of 34 millamperes, and the approximate coefficients $\mu = 200$, $r_p = 80{,}000$ ohms, and $g_m = 2500$ micromhos.

If the static grid voltage-plate current curves for a pentode or

for a beam-power tube (which behaves similarly) are plotted, and if dynamic curves for several load resistances are calculated, as explained in Fig. 179, dynamic curves of surprising shapes result. From the dynamic curves it is found that the load resistance that must be used is from about one-tenth to one-fifth the plate resist-

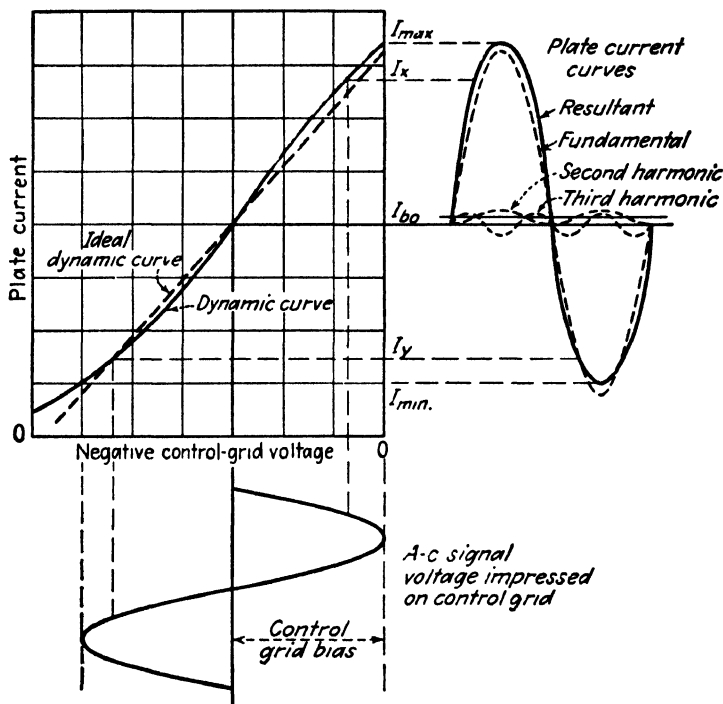



FIG. 186.—Showing the dynamic curve for a power pentode, and the nonlinear distortion that is caused. The dynamic curve here shown is for a load resistance about one-tenth the plate resistance. Higher load resistances give dynamic curves that are much more curved and that would cause greater nonlinear distortion.

ance or excessive distortion will result because the other dynamic curves are so crooked. The exact amount depends on the tube. A typical dynamic curve for a power pentode is shown in Fig. 186. This curve is for the tube and a load resistance equal to about one-tenth the plate resistance. For the pentode power tube previously considered, r_p = 80,000 ohms, but R_L should be *only 7000 ohms.* (Contrast this with a power triode where R_L is about twice r_p.) Similarly, a beam-power tube, the 6L6, has a plate resistance of 22,500 ohms, and for class A1 operation should use a load resistance of 2500 ohms.

Now such a great mismatch between the plate and load resistances means that much of the power-handling capacity of the tube is not utilized, but if this mismatch does not exist, the distortion will be excessive. In this connection it should be remembered that the equivalent circuit of a pentode is a generator of voltage μE_g and internal resistance r_p, and because it is working into a load resistance R_L it follows the usual power-transfer theory (page 99).

All these points would indicate that a triode would be superior to a pentode as a power-output tube. But it should be noted that the amplification factor is about 200 for a typical power pentode as compared with 3.5 for a comparable power triode. Although the power transfer with a pentode is poor, the amplification factor is so high that *much less* grid signal voltage is required with pentodes than with triodes for comparable outputs. In other words, in similar amplifiers, less voltage amplification is needed preceding power pentodes than preceding power triodes. This fact has caused pentodes and beam-power tubes to replace triodes to a large extent as audio-frequency power amplifiers, regardless of the fact that the nonlinear distortion often is greater than with triodes.

Pentodes and beam-power tubes are operated in push-pull circuits, but because the top of the dynamic curve "droops" downward, as shown in Fig. 186, instead of continuing to bend upward as does the dynamic curve of a triode (Fig. 179), the odd harmonic content in the output of pentodes is large. It will be recalled that only even harmonics are canceled in push-pull transformers.

It is possible to determine the power output, the percentage of distortion, and the proper load resistance from static curves much the same as for triodes.[1] If the dynamic curve is known, as in Fig. 186, the power output can be found for a single tube by the equation

$$P = [I_{\max} - I_{\min} + 1.41\,(I_x - I_y)]^2\,\frac{R_L}{32}. \qquad (104)$$

The percentage of second harmonic distortion can be found by the equation

$$\text{Second harmonic distortion} = \frac{I_{\max} + I_{\min} - 2I_{bo}}{I_{\max} - I_{\min} + 1.41(I_x - I_y)} \times 100. \qquad (105)$$

[1] RCA Receiving Tube Manual RC-14.

The percentage of third harmonic distortion can be found by the equation

$$\text{Third harmonic distortion} = \frac{I_{max} - I_{min} - 1.41\,(I_x - I_y)}{I_{max} - I_{min} + 1.41\,(I_x - I_y)} \times 100. \quad (106)$$

The percentage of total distortion is given by the equation

$$\text{Total distortion} = \sqrt{(\text{percentage 2d})^2 + (\text{percentage 3d})^2}. \quad (107)$$

If more than two harmonics are present, a third term is added in Eq. (107) in the same manner. This is a fundamental rule for combining various components of different frequencies. Thus the effective value of a nonsinusoidal wave equals the square root of the sum of the squares of the effective values of the components of different frequencies.

Audio-frequency Class B Power Amplifiers.—Operation in class B was discussed on page 322, where it was explained that the grid bias is approximately equal to the cutoff value, and the plate current is approximately zero with no applied alternating grid signal voltage. The plate current of a class B tube flows for approximately one-half of each cycle, as shown in Fig. 177, when the grid signal voltage is applied. Ordinary power triodes sometimes are used in audio-frequency class B power amplifiers, and of course these would require a large direct grid-bias voltage. More commonly, special tubes are used for audio-frequency class B power amplifiers, and these tubes are so constructed that the plate current essentially is zero with zero grid bias. In other words, they are so constructed that no bias is necessary to produce plate-current cutoff.

From Fig. 177 it is seen that the plate current flows in "pulses" as in a half-wave rectifier. This means that much nonlinear distortion exists and that the plate current contains the harmonic components listed on page 221 when a sinusoidal voltage is impressed on the grid. For *audio-frequency* power amplifiers, *two tubes* in a push-pull circuit similar to Fig. 183 are used. As explained for the push-pull circuit, the direct-current components (which are small because the tubes operate near plate-current cutoff) flow in opposite directions through the identical halves of the primary of the output transformer and their magnetizing effects cancel. Since in the push-pull circuit the control grids are driven alternately pos-

itive and negative by the impressed signal, and since the tubes operate near cutoff, one tube passes current, and then the other. Thus the statement is made often that one tube passes one half of an impressed sinusoidal signal, the other tube passes the other half, and as a result the full cycle of signal is amplified.

A more exact analysis discloses that for two tubes in an audio-frequency push-pull class B power amplifier, a diagram somewhat like that of Fig. 184 applies. The analysis indicates that, as in the case Fig. 184, the even harmonics cancel and the odd harmonics (and the fundamental) add. In this connection it should be remembered that for a wave such as Fig. 177 the various components are acting at all times, and that they total zero during the period that the current is zero. Thus with two tubes in an audio-frequency class B power amplifier the even-harmonic nonlinear distortion is canceled in the primary circuit of the output transformer. The fundamental components (which are, as explained on page 221, the first alternating-current components of half waves such as Fig. 177) add to produce a strong output signal.

As mentioned earlier in this section, special tubes that need no biasing voltage to produce plate-current cutoff are used generally in class B audio power amplifiers. Often these are special tetrodes or pentodes that are connected *as triodes*. The characteristics of two such tubes in push-pull are shown in Fig. 187. Note that but little plate current flows when the signal is zero. When current does flow, it is largely useful signal current. This means that the plate-circuit efficiency of a class B power amplifier is higher than that of a comparable class A power amplifier. In class A power amplifiers a large value of direct current flows at all times, and this component contributes nothing to the signal.

If two tubes operate as in Fig. 187, they will be in class B2 because grid current will flow. In fact most audio amplifiers of this kind operate with grid current flowing. During operation the tube preceding the class B2 stage must supply alternating-current power to the grid circuit. This means that the stage preceding must be a power amplifier of comparatively small power capacity, and that the transformer between the two stages (the class B2 input transformer) must be able to handle the power without distortion. Such transformers are called **driver transformers** and differ from interstage transformers because they pass power between the two stages and usually have fewer secondary than

primary turns. As an illustration, two type 46 tetrodes, connected as triodes with the two grids tied together and operating in class B2, require a grid-driving power of 650 milliwatts and have an output of 20 watts.

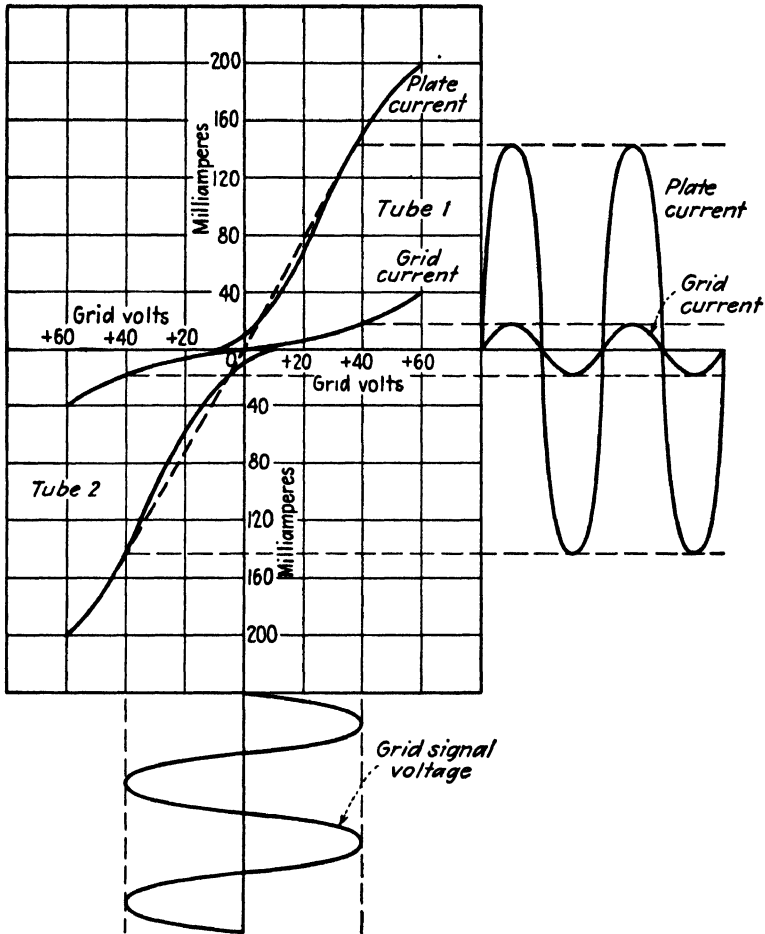
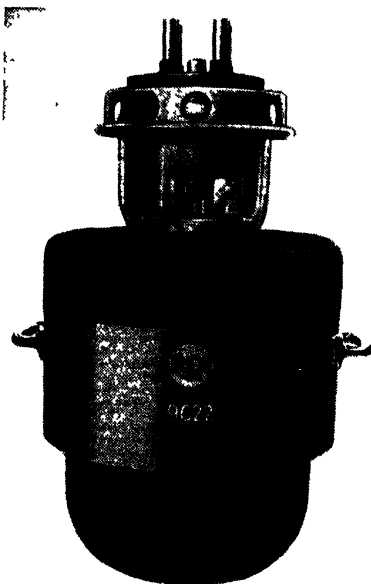



Fig. 187.—Showing the plate and grid characteristics of pentodes that operate with zero grid bias. Plate-current curves are dynamic curves. When a grid signal voltage is impressed, there will be some distortion because the dynamic curves are not straight lines.

Class B2 amplifiers are more efficient, are lighter, and have other advantages, such as low cost and compactness, over comparable class A audio amplifiers. On the other hand, distortion is rather high with class B2 amplifiers. They are well suited as audio-

frequency power amplifiers in sound-amplifying systems, such as the ones used in stadiums, etc. It is common practice to speak of class B2 audio-frequency power amplifiers merely as class B audio-frequency amplifiers.

Radio-frequency Power Amplifiers.—The purpose of the radio-frequency power amplifier is to furnish the radio-frequency power

A large power-output triode. The plate is inserted in cooling fins through which air is forced at the rate of about 1800 cubic feet per minute. The over-all height is about 25 inches. One use of this tube is as an amplifier in an amplitude-modulated radio transmitter. For such purposes maximum operating conditions are direct plate voltage 12,500 volts, direct grid voltage −2000 volts, maximum direct plate current 4 amperes, plate input power 50 kilowatts, plate dissipation 14 kilowatts, signal power output 38 kilowatts. This tube has a multistrand tungsten filament, requiring a heating current of 415 amperes at 19.5 volts. (*Radio Corporation of America.*)

for driving the transmitting antenna. Another, and increasing use, is to furnish high-frequency power for industrial purposes, such as heating by high-frequency magnetic or electric fields. A radio-frequency power amplifier may be operating over a band of frequencies as when it is amplifying the carrier and the two sidebands in an amplitude-modulation radio system. Or a radio-frequency power amplifier may be operating at a single frequency as in industrial-heating apparatus.

Many radio-frequency power amplifiers use power-output triodes. This is particularly true for the large amplifiers with outputs from several hundred watts up to many kilowatts. In small power amplifiers, and in those of intermediate sizes (up to several hundred watts), power tetrodes and particularly power pentodes are used. Here again the beam-power tube is considered to be a pentode because it performs so much like a pentode. The power-handling capacities of tetrodes and pentodes gradually is being extended upward.

Within recent years many developments of importance have been made, especially at the very high and ultrahigh radio frequencies (page 372). The discussions in this chapter must, of necessity, be

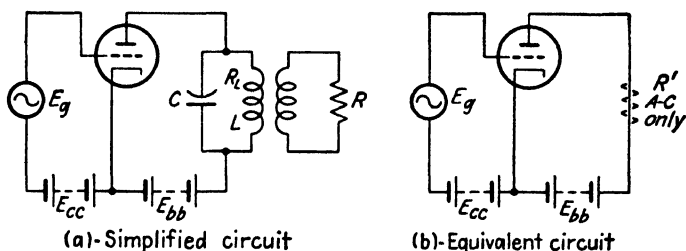



FIG. 188.—Basic circuits of a radio-frequency class C power amplifier.

limited to the radio-frequency power amplifiers that operate over the lower radio frequencies in the vicinity of from 500,000 to 30,000,000 cycles.

Radio-frequency Class C Triode Power Amplifier.—Logically, the class B radio-frequency power amplifier should be considered before the class C type. This is not being done because the class C type is of more importance, and if it is understood, the operation of the class B type is easily explained.

The basic circuit of a radio-frequency class C power amplifier is shown in Fig. 188. The tube is assumed to be feeding signal power to the load resistance *R*, which can be thought of as representing the input resistance of a radio-transmitting antenna. This is coupled into the plate circuit of the tube through the coupling network (page 542). When this is tuned to resonance, the load in the plate circuit is an equivalent resistance *R'*. This is an alternating-current resistance and has been shown as a broken line. It is assumed that the direct-current resistance is zero.

In the design of a radio-frequency power amplifier, efficiency is

of importance. Also, for economic reasons it is generally desired to obtain the required output from the tube of the smallest size. For purposes of design, constant-current curves are very convenient, and will be used for illustrating the method, which is one of several possible schemes. As will be observed, the design is a cut-and-try process, the number of attempts depending on the experience of the designer.

Illustrative Problem.—A small radio-telegraph transmitting set is to have 50 watts output at 2.6 megacycles, and is to be operated from an existing 750-volt power supply. Select the tube for this purpose, and make the calculations necessary to determine its power output and efficiency for a given set of conditions. The general procedure is to decide on a certain set of tube operating conditions, and if these prove satisfactory, design the circuit in such a manner that the tube operates according to the assumed conditions.

Solution.—Step 1. Fifty watts is a small amount of power, and the examination of manufacturers' catalogues indicates that the type 809 tube probably will be satisfactory. Constant-current curves for this tube are shown in Fig. 189.

Step 2. It is decided to try operation at three times plate-current cutoff (page 184) and at 750 volts on the plate, using the existing power supply (and assuming that the voltage drop in the output coil is zero). This fixes the Q **point** on Fig. 189, as indicated, and the negative grid-bias voltage is -55 volts.

Step 3. The next step is to decide to what maximum positive value the grid voltage will rise, and to what minimum value the plate voltage will fall. These are important points. When the signal voltage drives the grid positive, a large current flows in the plate circuit, and the plate voltage falls (page 274). If the plate voltage falls to a value that approaches, or falls below, the voltage of the grid, then the grid may draw a large current that otherwise would go to the plate, thus damaging the tube. It is decided that the grid voltage will be permitted to rise to $+80$ volts and that the plate voltage will be allowed to fall to 100 volts. This fixes the **T point** of Fig. 189.

Step 4. With the Q point and T point known, the grid-voltage swing and resulting plate-voltage variations can be found. The Q point is the quiescent or no-signal condition, and as such the direct-plate voltage is 750 volts. But if the plate voltage falls to 100 volts when the grid voltage is applied, then the peak value of the plate voltage is $750 - 100 = 650$ volts. With no signal on the grid, it is 55 volts negative with respect to the cathode, and for operation at the assumed T point the positive half cycle of the grid signal-voltage swing must be $55 + 80 = 135$ volts. The corresponding effective values of the alternating-current components are $E_p = 0.707 \times 650 = 459$ volts and $E_g = 0.707 \times 135 = 95.5$ volts. These voltage variations are shown in Fig. 189. The grid signal voltage is furnished by the preceding stage that drives the power-amplifying tube under consideration. This grid-voltage swing causes the plate-current

variations of Fig. 178. These pulsations of plate current contain an average, or direct-current, value and various alternating-current components. As in the case of the half wave (page 221), the lowest alternating-current component has the same frequency as the signal impressed on the grid. The direct-current flows through the coil in the plate circuit with negligible voltage drop, and essentially the entire plate-supply voltage

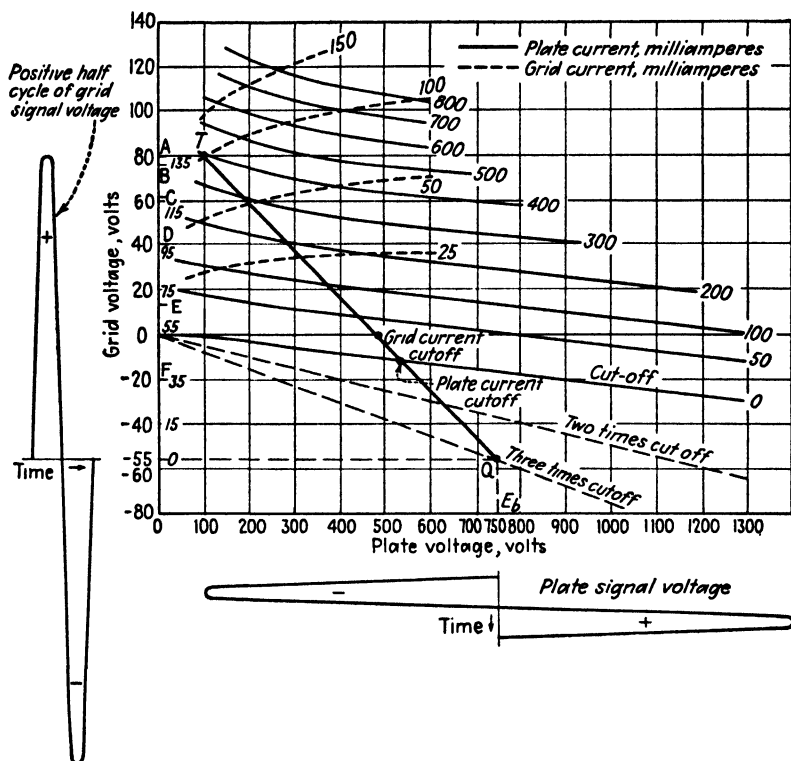



FIG. 189.—Curves for studying the design of a radio-frequency class C amplifier. The curves are for a type 809 triode, a small power-output tube.

E_{bb} is on the plate. The tuned parallel circuit and inductively coupled load offer an equivalent resistance R' to the fundamental component of the pulsating plate current. There is an IR' drop caused by the fundamental component. When this is positive, it adds to the voltage E_{bb}, causing the plate voltage to rise above the value (essentially) E_{bb}. For Fig. 189 this maximum value is +1400 volts. When this drop is negative, it subtracts from the value E_{bb}, causing the plate voltage to fall to +100 volts. Negligible voltage drop occurs in the plate circuit for the harmonic components of the plate current, because this circuit is not in parallel resonance to the harmonic frequencies, and, hence, offers little impedance to them.

Step 5. It is now necessary to evaluate the currents that will flow.[1] The average current (Fig. 178) is found by the equation

$$I_{av} = 0.0833\ (05A + B + C + D + E + F). \qquad (108)$$

The maximum value of the fundamental component of the current (Fig. 178) is

$$I_{max} = 0.0833\ (A + 1.93B + 1.73C + 1.41D + E + 0.052F). \qquad (109)$$

In these equations A, B, C, D, E, and F are the total instantaneous values of the plate and grid currents that flow at each 15° increment of grid-driving voltage. To find these it is first necessary to find the instantaneous grid voltage at each point A, B, etc. Thus, $A = E_{g(max)} \times \cos\theta = 135 \cos\theta = 135 \times 1 = 135$ volts. Point A is marked on Fig. 189. Point B is at $B = E_{g(max)} \times \cos\theta = 135 \cos 15° = 135 \times 0.966 = 130$ volts. This locates point B on Fig. 189. Likewise, $C = 117$ volts, $D = 95$ volts, $E = 68$ volts, and $F = 35$ volts. The plate currents and grid currents that correspond to these voltages are then obtained from the curves of Fig. 189 following the T-Q line. Thus at point A, $i_b = 0.40$ ampere, and $i_c = 0.10$ ampere; at point B, $i_b = 0.38$ ampere, and $i_c = 0.08$ ampere; at point C, $i_b = 0.31$ ampere, and $i_c = 0.06$ ampere; at point D, $i_b = 0.21$, and $i_c = 0.03$; at E, $i_b = 0.07$, and $i_c = 0.01$; at F, $i_b = 0$, and $i_c = 0$. From Eqs. (108) and (109) the following result:

Grid currents

$$I_{av} = 0.0833(0.5 \times 0.10 + 0.08 + 0.06 + 0.03 + 0.01 + 0) = 0.0192 \text{ ampere.}$$

$$I_{max} = 0.0833(0.1 + 1.93 \times 0.08 + 1.73 \times 0.06 + 1.41 \times 0.03 + 0.01 + 0.052 \times 0) = 0.0342 \text{ ampere.}$$

$I_g = 0.707 \times 0.0342 = 0.0242$ ampere, effective value.

Plate currents

$$I_{av} = 0.0833(0.5 \times 0.40 + 0.38 + 0.31 + 0.21 + 0.07 + 0) = 0.0975 \text{ ampere.}$$

$$I_{max} = 0.0833(0.4 + 1.93 \times 0.38 + 1.73 \times 0.31 + 1.41 \times 0.21 + 0.07 + 0.052 \times 0) = 0.17 \text{ ampere.}$$

$I_p = 0.707 \times 0.17 = 0.120$ ampere, effective value.

Step 6. The grid-driving power and apparent grid dissipation can be computed now. This is the alternating-current power that the driving stage must furnish to the grid of the class C amplifier. This will be approximately the effective value of the grid signal voltage (Step 4) multiplied by the effective value of the alternating grid current (Step 5), or

$$\text{Grid-driving power} = 95.5 \times 0.0242 = 2.31 \text{ watts.}$$

When this power is put into the grid circuit, this circuit acts like a rectifier, producing a direct current of $I_{av} = 0.0192$ ampere. This flows through the source of grid-bias voltage E_{cc} of Fig. 188 in a direction such that it supplies power to this source. If E_{cc} is a battery, it "charges" the battery.

[1] See Chaffee, E. L., A Simplified Harmonic Analysis, *Review of Scientific Instruments*, October, 1936.

The power furnished in this instance is the product of the current and the biasing voltage or

Grid-battery charging power = 55 × 0.0192 = 1.06 watts.

The remainder of the power put into the grid circuit will be dissipated inside the tube, thus

Apparent grid dissipation = 2.31 − 1.06 = 1.25 watts.

Step 7. Calculate the input power to the plate circuit of the tube. This will be the product of the direct-plate voltage and current from Step 5.

Plate-input power = 750 × 0.0975 = 73 watts.

Step 8. Calculate the alternating signal-power output delivered by the power-amplifier tube to the load. This will be the product of the alternating signal-voltage drop across the equivalent load resistance (Step 4) and the effective value of the signal current, or fundamental component (Step 5), through the equivalent load resistance.

Signal-power output = 459 × 0.12 = 55 watts.

Step 9. Calculate the plate efficiency of the power tube as operated under the conditions specified. This will be the alternating signal-power output (Step 8) divided by the direct-current power input (Step 9).

$$\text{Plate efficiency} = \frac{55}{73} = 0.753 \times 100 = 75.3 \text{ per cent.}$$

Step 10. Calculate the power amplification. This will equal the signal-power output (Step 8) divided by the signal-power input (Step 6).

$$\text{Power amplification} = \frac{55}{2.3} = 23.9.$$

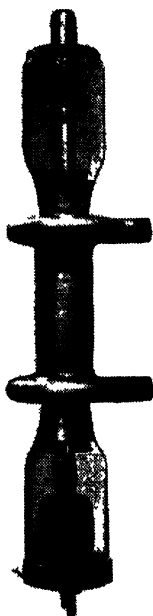
Step 11. Calculate the equivalent resistance R' of Fig. 188, which the parallel circuit with the coupled load resistance R must offer to the tube. There are several ways of obtaining this, one being to divide the effective value of the voltage drop (Step 4) by the effective value of the current (Step 5).

$$\text{Equivalent plate-load resistance} = \frac{459}{0.12} = 3820 \text{ ohms.}$$

Step 12. Calculate the inductance and capacitance that the parallel circuit of Fig. 188 must have in order to give the correct equivalent load resistance. To do this, the effective Q (page 49) of the parallel load circuit must be assumed. This often is taken at about 12. If the Q is much larger, the circuit will be too sharpely tuned, and if it is much lower the tuning will be too broad. Then, from Eq.(24), page 84, at a frequency of 2.6 megacycles (the frequency at which the tube is being operated), $R_e = \omega L Q_e$, and $L = R_e/(\omega Q_e)$.

$$\text{Self-inductance of load circuit} = \frac{3820}{2 \times 3.14 \times 2.6 \times 10^6 \times 12}$$
$$= 19.5 \times 10^{-6} \text{ henry.}$$

The required capacitance to produce resonance in the plate circuit can be found approximately from the usual resonance equation (page 75).

$$\text{Capacitance of load circuit} = \frac{1}{(2 \times 3.14 \times 2.6 \times 10^6)^2 \times 20.0 \times 10^{-6}} = 192 \times 10^{-12} \text{ farad.}$$

A large water-cooled power-output triode. The grid connections are at the top, the plate connections at the center, and the filament connections at the bottom. The plate is surrounded with a water jacket through which water circulates at the rate of not less than 3 gallons per minute. The water pipe connections are shown at the upper and lower ends of the center structure. The height is about 25 inches. One use of this tube is as an amplifier in amplitude-modulated radio transmitters of large capacity. For such purposes the direct plate voltage may be as high as 12,000 volts, the grid bias −600 volts, the direct plate current 1.5 amperes, the plate dissipation 6 kilowatts, and the signal power output 12 kilowatts. This tube has a tungsten filament that requires a heating current of 41 amperes at 21.5 volts. (*Western Electric Co.*)

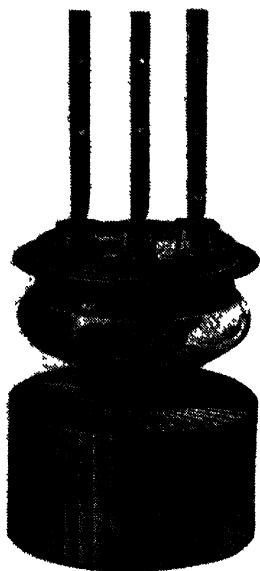
Further calculations can be made as desired to determine the value of the mutual inductance required to couple the resistance of the load into the plate circuit so that the conditions of Step 11 are satisfied. In this example the calculated power output of 55 watts is very close to the 50 watts desired. If it had not been, a new set of operation conditions (T and Q points) would be assumed, and the calculations repeated. It is recognized that these calculations are very approximate, but it should be remembered also that in the operation of a power amplifier, circuit adjustments are made to obtain correct operation.

Uses of Class C Power Amplifiers. With this amplifier the plate current flows in pulses for time intervals considerably less than one-half cycle, that is, for less than 180° of the cycle. For the amplifier of the preceding section, the angle of flow can be found as follows: Plate current starts to flow when the grid is less negative than the cutoff point. From Fig. 189, this occurs at a grid voltage of about 42 volts. Thus, when the alternating grid signal voltage has gone from 0 to 42 volts toward a maximum of 135 volts, grid current flows. If a sinusoidal grid voltage is assumed, the sine of the angle at which plate current flows is $\sin \theta = 42/135 = 0.311$, and $\theta = 18°$, approximately. Plate cur-

rent flows until the angle is 162° and then stops flowing. Thus the angle of plate-current flow is $\theta_b = 144°$ (Fig. 178). For the grid circuit, current will flow when the grid is driven positive, and for the tube under consideration this occurs at +55 volts for the grid-driving signal. The angle at which grid current starts is $\sin \theta = 55/135 = 0.407$ and $\theta = 24°$. As explained for the plate current, the angle of grid-current flow is $\theta_c = 132°$. Sometimes the angle of flow is specified as one-half the angle given here.

Because the angle of flow of plate current is small, the plate current is distorted badly. This means that many frequency components exist in the output. If a single frequency is to be amplified, a class C amplifier is usually employed. The tuned parallel circuit offers resistance at the frequency to which it is tuned, and draws power from the tube at this frequency. At all other frequencies, the plate load is reactive and draws little power at these harmonics. Here it is well to recall that at radio frequencies a harmonic is many cycles away from the fundamental. Thus for the amplifier of the preceding section, if the fundamental is 2.6 megacycles, or 2,600,000 cycles, the second harmonic will be at 5,200,000 cycles, and the third harmonic is at 7,800,000 cycles. For these reasons, it is practicable to amplify (and distort) a single frequency, select the fundamental, and reject the harmonics.

Sometimes it is desired to increase the *frequency*, and this may be done by tuning the parallel circuit in the plate lead to one of the harmonics *instead* of the fundamental. Thus if the output circuit is tuned to 5.2 megacycles, then this frequency will be amplified and passed on, and all others (including the fundamental) are rejected. The device then is a **frequency doubler.** Or if the output circuit is tuned to 7.8 megacycles, the device then becomes a **frequency tripler.** By the use of such **frequency multipliers,** frequencies can be successively stepped up to high values. Of course if the frequency is to be increased by selecting a specific harmonic, the tube should be operated so that the desired harmonic is large. Different angles of plate-current flow accentuate certain frequencies and minimize others. An approximate rule to follow is that for frequency multiplication the angle of flow should be

$$\theta_b = \frac{180°}{n}, \qquad (110)$$

An air-cooled triode of the type used at very high radio frequencies in frequency-modulated transmitters. The plate is inserted in cooling fins through which air is forced at a rate of about 275 cubic feet per minute. The over-all height is about 11 inches. The filament connections are at the top, and the grid connection is the metal flange immediately below the filament leads. One use of this tube is as a grounded-grid class C amplifier in frequency-modulated transmitters. Typical operation conditions are direct plate voltage 4000 volts, direct grid bias −350 volts, and direct plate current 0.8 ampere. The signal power output is from about 1 to 2 kilowatts, depending on the type of service, frequency, etc. This tube has a thoriated-tungsten filament, requiring 29 amperes at 12.6 volts. (*Radio Corporation of America.*)

where n is the desired output frequency divided by the grid-driving frequency. Thus for a frequency tripler $n = 3$, and $\theta_b = 60°$. The power outout of a frequency multiplier is about $1/n$ times that of the corresponding class C amplifier. Thus if a class C amplifier gives an output of 50 watts at the fundamental frequency, about 15 watts could be obtained at the third harmonic.

Class C amplifiers are not used to amplify an amplitude-modulated carrier composed of a carrier frequency and the two sidebands. Apparently the distortion produced by a class C amplifier is too great for such purposes.

Radio-frequency Class B Triode Power Amplifiers.—These are used extensively as power-output stages in radio transmitters to amplify signals composed of the carrier and the sidebands (page 308). The basic circuit and the equivalent circuit are as in Fig. 188. The tube is biased to cutoff, and plate current flows for 180°. Of course the plate current will be pulses (Fig. 177), and the desired signal components are selected by the tuned plate circuit, as in the class C amplifier. The analysis of the class B amplifier is simimilar to that of the class C amplifier.

The class B amplifier is used to increase the radio-frequency signal power of signals composed of a carrier and side-bands. It often is called a **linear power amplifier.** It should be remembered, however, that the plate current is badly distorted, and that harmonics are created because of this nonlinear distortion. The tuned plate-load circuit selects the signal frequencies desired, as explained in the

preceding section. Because plate current flows for 180°, the distortion is less than for class C operation where the angle of flow is less than 180°. In fact, if a sinusoidal wave is impressed on a class B tube, the output current will be the same as that of a half-wave rectifier (page 221). A fundamental difference exists, however. A two-electrode rectifier tube does not add energy to the circuit, but a three-electrode class B tube adds energy from the plate power supply.

The tuned load circuit of Fig. 188 often is referred to as a **tank circuit,** and is discussed from the energy-storage standpoint. Such analyses often do not distinguish between the theory of self-oscillating circuits and driven circuits. The class C or the class B circuits of Fig. 188 are not self-oscillating, but are driven by the source of grid signal voltage and the tube. As such, ordinary circuit theory, as employed in the preceding pages, is sufficient to explain their operation.

Vacuum-tube Voltmeters.—In the operation and adjustment of class B or class C radio-frequency power amplifiers the grid bias can be measured with an ordinary direct-current voltmeter. The source of the bias can be a power supply, or a radio "B" battery. Because of the charging action (Step 6, page 346) such a battery lasts for a long time. To operate at the selected T point, the grid-driving voltage must be known. Of course this can be measured with one of the many commercially available vacuum-tube voltmeters.

For operation at the selected Q point, the direct-plate voltage must be known. This can be measured with a direct-current voltmeter of sufficient range. The adjustment of the tuned plate circuit will greatly affect the T point, and a vacuum-tube voltmeter can be used to determine this point. It will be remembered that in the calculations the T point is selected arbitrarily, but the circuit adjustments fix this point in actual operation. To be more specific, Step 11, page 347, shows that the equivalent plate-load resistance should be 3920 ohms. This value will give correct operation. The voltage drop across it will be such that if the grid voltage is driven to a peak of 135 volts, the plate voltage will fall to a minimum of 100 volts. But this value of 3920 ohms is a fictitious alternating-current resistance, and the tuned circuit and coupled circuit must be so adjusted that R of Fig. 188 appears as an R' of 3920 ohms.

If commercial vacuum-tube voltmeters are not available, or if their frequency range or maximum readings are not satisfactory, then simple vacuum-tube voltmeters can be constructed to determine if the grid-voltage swing and plate-voltage swing are such that operation is at the selected T point. If it is not, then operation does not represent calculated conditions.

The Peak Vacuum-tube Voltmeter.—The circuit of a voltmeter that will measure the peak of the grid signal voltage applied to the grid of a power-amplifying tube is shown in Fig. 190 connected

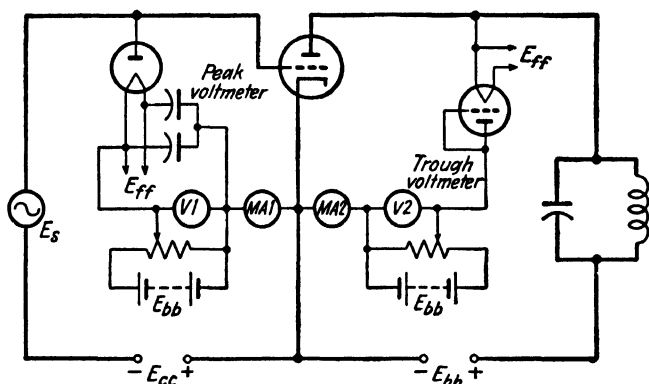



FIG. 190.—Circuit arrangement for studying a class C amplifier. The peak voltmeter is used to measure the peak value of the grid-signal voltage. The trough voltmeter is used to measure the minimum voltage to which the plate falls when the tube acts as a class C power amplifier.

across the grid circuit. It will be noticed that the voltage divider *in the cathode circuit biases the cathode* of the voltmeter tube positively with respect to the plate. The plate of this tube will not take electrons from the cathode, until the plate is positive with respect to the cathode; also milliammeter $MA1$ will not indicate until this happens. Now at a given instant, the positive half cycle of the grid-signal swing is attempting to drive the plate of the vacuum-tube voltmeter tube positive. It is opposed in this action by two potentials: (*a*) that of the voltage divider and (*b*) that of the grid-bias source. Thus when the voltage divider is adjusted so that milliammeter $MA1$ *just reads*, then the peak of the voltage wave will equal, approximately, the sum of the bias voltage E_{cc} and the voltage read by the voltmeter $V1$ connected across the voltage divider. For the class C amplifier of the preceding pages, where it is desired to have a peak grid voltage of 135

volts to drive the grid to the assumed T point, voltmeter $V1$ should read $135 - 55 = 80$ volts when $MA1$ first indicates, assuming of course sinusoidal grid excitation.

The Trough Vacuum-tube Voltmeter.—The plate voltage of a power amplifier will rise and fall as shown by Fig. 189. Thus it falls to a minimum value, or *trough* as it is called. This is at the T point, as explained in the preceding pages. The vacuum-tube voltmeter connected across the plate circuit of Fig. 190 can be used to determine this minimum, or trough, value to which the plate voltage falls during each cycle of operation. It will be noted from Fig. 190 that the plate-power supply E_{bb} tends to keep the *cathode* of the vacuum-tube voltmeter tube positive with respect to the plate. The voltage divider in the plate circuit of this tube tends to keep the *plate* positive. The milliammeter $MA2$ will read only when the plate is more positive than the cathode is positive. Thus if the voltage divider is set so that milliammeter $MA2$ just indicates, it indicates because the negative half cycles of the alternating signal-voltage drop in the plate circuit of the power-amplifier tube have driven its plate and the cathode of the trough voltmeter tube down essentially to the same potential as indicated by voltmeter $V2$. Thus, if $V2$ reads 100 volts, the trough of the plate alternating signal-voltage swing is 100 volts, giving the correct T point for the example on page 344. In Fig. 190, a diode is shown in the grid-circuit peak voltmeter, and a triode, connected as a diode, is shown in the plate-circuit trough voltmeter. There is no reason why this specific combination must be used. The tubes and the voltmeter circuits should contain little capacitance and inductance, and the tubes, sockets, etc., must be able to stand the voltages encountered. These circuits are adaptations of a type of vacuum-tube voltmeter known as the **slide-back voltmeter.**

The Slide-back Vacuum-tube Voltmeter.—The two vacuum-tube voltmeters just described are designed for measuring peaks and troughs in radio-frequency power amplifiers. The vacuum-tube voltmeter shown in Fig. 191 is useful for measuring alternating voltages in almost any circuit. The condenser and resistor in the grid circuit are for isolating the tube from direct-current components that may be present in the signal voltage to be measured. A good paper condenser which has a capacitance of the order of 0.05 microfarad and which will stand several hundred volts is

satisfactory. The resistor in the grid circuit is of the order of 0.5 megohm. The condenser across the milliammeter may be similar to the input condenser. The milliammeter should be very sensitive, perhaps requiring 1 milliampere for full-scale deflection. With no signal voltage applied, the voltage divider is adjusted for

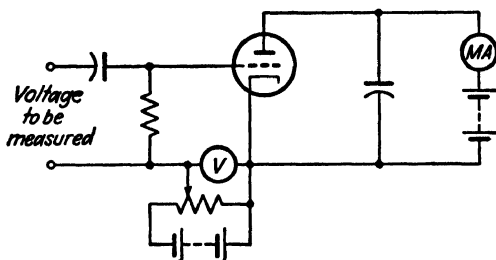



FIG. 191.—A slide-back vacuum-tube voltmeter.

a very small reading of the milliammeter, and the indication of the voltmeter is noted. Then, the signal voltage to be measured is applied, and the voltage divider is adjusted until the reading of the milliammeter is the same as before. The difference between the two voltmeter indications is approximately the peak value of the voltage to be measured. A tube with a sharp cutoff should

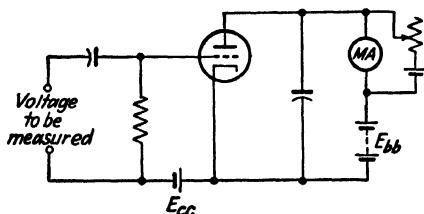



FIG. 192.—Basic circuit of a rectifier-type vacuum-tube voltmeter using a triode. The exact method of operation depends on the magnitude of the grid bias as shown in Fig. 193.

be used. Sometimes a tetrode or a pentode is employed, in which event the screen grid is used as a control grid, and the usual control grid is connected to the cathode. This is done to obtain sharp cutoff. There is an error involved with this slide-back voltmeter, but with large signals and tubes with sharp cutoff the error is small. Calibration curves and correction curves can be used where high accuracy is needed.

The Rectifier-type Vacuum-tube Voltmeter.—There are many variations of the basic circuit arrangement shown in Fig. 192. The input condenser and resistor are for isolating the tube from

direct voltages that may be present in the input voltage to be measured. These were discussed in the preceding paragraph. The signal voltage to be measured is impressed where indicated, and because of this applied grid signal voltage a change in plate current occurs. This is measured by the milliammeter in the plate circuit. From calibration curves made by applying known voltages, the magnitude of the unknown voltage can be determined. The zero balance circuit is for adjusting the initial reading,

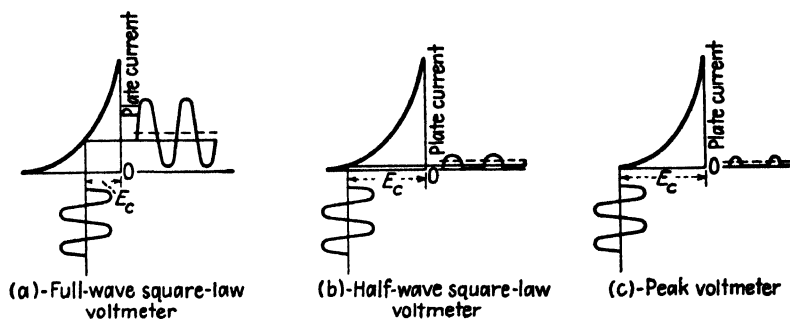



FIG. 193.—Methods of operating the voltmeter circuit of Fig. 192. These methods of operation are determined by varying the grid bias E_c. This is accomplished readily if a voltage divider is substituted for E_{cc} of Fig. 192. When no signal voltage is impressed on the input, the direct plate current is as indicated by the distance from the X axis upward to the solid line. In (c) this essentially is zero. With the signal voltage to be measured impressed on the grid, the *increase* in direct plate current is as indicated by the distance between the solid line and the broken line. For (a) the term "square-law" means that with operation as shown the *change* in plate current is proportional to the *square* of the *effective* value of the signal to be measured. For (b) the term "square-law" means that the *change* in plate current will be proportional to the *square* of the *effective* value of the positive half cycle. For (c) plate current flows only on the peaks of the positive half cycle.

with no applied signal, to zero. Normal plate current with no applied signal voltage will be flowing *up* through the milliammeter. With the rheostat and battery enough current is forced *down* through the milliammeter so that the net current is zero. Of course, this is just an approximate explanation; actually, the circuit is adjusted until the milliammeter terminals are at zero potential, and hence no current flows through it. When adjusted, the milliammeter indicates only the *change* in current caused by the voltage to be measured. These vacuum-tube voltmeters are of three basic types, as shown in Fig. 193, depending on the magnitude of the grid bias. These voltmeters can be calibrated with known voltages and at various frequencies.

Other rectifier-type vacuum-tube voltmeters employ small

diodes, often mounted in a metal housing or probe. These diodes rectify the impressed signal voltage, and transmit only a direct-current component down the connecting cable leads to the voltmeter circuit. In this way, very small input capacitances of a few micromicrofarads are obtainable, and the voltmeters may be used up to hundreds of megacycles.

Measurements of Power at Radio Frequencies.—Although vacuum-tube voltmeters have been available for many years, there is no corresponding simple and rugged wattmeter for high-frequency circuits. Vacuum-tube wattmeters, and other special radio-frequency power-measuring devices, have been developed, but have never achieved wide usage.

Current readings are used widely for determining power at radio frequencies. The radio-frequency current flowing into a circuit is measured with a thermocouple ammeter, and the effective resistance of the circuit is measured with a bridge; then the power dissipated in the circuit is $P = I^2R$. Of course this method depends on the accurate measurement of the current and the resistance. Thermocouples are available which have little error up to about 100 megacycles, and bridges are available commercially by which radio-frequency measurements can be made to about this frequency.

Voltage readings are used extensively for determining power at radio frequencies. The radio-frequency voltage drop across a circuit is measured with a vacuum-tube voltmeter, and the resistance is measured with a radio-frequency bridge. The power dissipated in the circuit is then $P = E^2/R$.

Calorimetric methods are used sometimes for measuring the power dissipated. Here the radio-frequency device is enclosed so that the temperature rise can be measured and the power loss computed. A variation of this method often is applied to large water-cooled tubes. From the temperature rise of the circulated cooling water, the power loss can be computed. The total power input minus the power lost will give the useful power delivered to the radio circuit.

Neutralization of Radio-frequency Power Amplifiers.—Triodes are used extensively in radio-frequency power amplifiers of all sizes. Tetrodes and particularly pentodes are used sometimes in small amplifiers and in those of intermediate size, but for power

amplifiers of large capacity (that deliver many kilowatts) triodes are employed.

The input impedance of triodes was considered briefly on page 287, where it was explained that under certain conditions there was a tendency for triodes to feed back to the grid circuit an amplified signal from the plate circuit. This may cause a tube to oscillate, and in power amplifiers this is undesired because it is the purpose of an amplifier to operate at the signal frequency impressed on the grid, and not to generate a signal frequency of its own. In radio-frequency power amplifiers using triodes, special precautions must

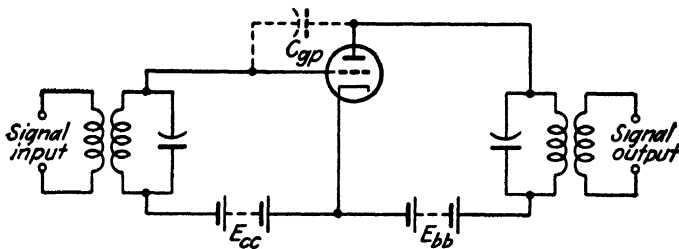



FIG. 194.—A tuned-plate tuned-grid radio-frequency amplifier, showing the inter-electrode capacitance between the grid and plate as an external condenser C_{gp}. With triodes this capacitance causes signal feedback and may cause oscillations.

be taken to ensure that the tubes do not oscillate, and this is known as **neutralization.**

Reason for Signal Feedback.—A simplified circuit of a radio-frequency power amplifier is shown in Fig. 194. The grid-plate capacitance that exists *within* the tube is shown as a condenser C_{gp} connected externally. Both the input and output circuits are tuned to resonance and are equivalent to resistance at the frequency of resonance.

For an amplifier tube to oscillate, the amplified signal in the plate circuit must feed back signal to the grid circuit in such a way that the impressed signal and the feedback signal add to increase the signal voltage actually impressed between grid and cathode. Under this condition, the action is cumulative; the initial signal is amplified and fed back, and then causes a larger amplified signal and increased feedback, and so on.

Theoretically a circuit with resistance in the grid and plate circuits (a resistance-coupled amplifier) should not oscillate, and hence Fig. 194 with tuned circuits (equivalent to resistance) should

not oscillate. As mentioned on page 290, triodes tend to oscillate when the load in the plate circuit is inductive. But the tuning of a circuit such as Fig. 194 is not perfect, and although the tube is supposed to operate only at one frequency or only over a narrow band of frequencies, it will tend to oscillate at other frequencies. There will be sufficient feedback to the grid circuit of signal of the proper phase to cause oscillations, and this feedback occurs through the grid-plate interelectrode capacitance C_{gp}.

The obvious ways to neutralize a radio-frequency power amplifier and to prevent oscillations from occurring are (*a*) prevent the feedback, and (*b*) intentionally feed back enough signal of the proper phase to neutralize or offset the effect of the signal unintentionally fed back through the grid-plate capacitance.

Coil Neutralization.—This is a very simple method of neutralization, and is accomplished by connecting a small coil and condenser in series between the plate and grid of the tube of Fig. 194. The condenser is necessary to isolate the direct voltages of the grid and plate. The sizes of the coil and condenser are such that at the frequency at which the amplifier is to operate, the combined *series* reactance of the two is *inductive.* This inductive reactance is in *parallel* with that of condenser C_{gp}. The circuit is then adjusted so that parallel resonance occurs between the two branches, one composed of C_{gp}, and the other of the neutralizing coil and condenser, and at resonance a parallel circuit has very high impedance. By this simple means, the impedance of the path from the plate circuit to the grid circuit in a triode is made very high, and appreciable signal feedback cannot occur. It should be noted that this neutralization is effective at one frequency only and that it is best suited for power amplifiers of fixed operating frequency. An alternate explanation is that the coil and series condenser adjusted so that they are *equivalent to inductance,* feed back a 90° lagging current *from the same point* that the interelectrode capacitance C_{gp} feeds back a 90° leading current. These two currents would cause opposing voltage drops in the grid circuit, and no resultant signal feed-back voltage would exist.

Condenser Neutralization.—Attention again is called to the fact that with coil neutralization current is fed back from the *same point* that feedback occurs through the interelectrode capacitance C_{gp}. The coil, allowing a lagging current to flow, neutralizes the effect of the leading current passed by the plate to grid capacitance.

It is very convenient to use a small variable air capacitor to produce neutralization. If such a condenser were connected from the plate to cathode (as was the coil of the preceding paragraph), then this "neutralizing" condenser would not produce neutralization at all; instead, it would feed back more of the same type of current, and would encourage oscillations. If a condenser is to be used for producing neutralization, then it must be connected to a point where the signal voltage is 180° out of phase with the signal voltage across C_{gp}.

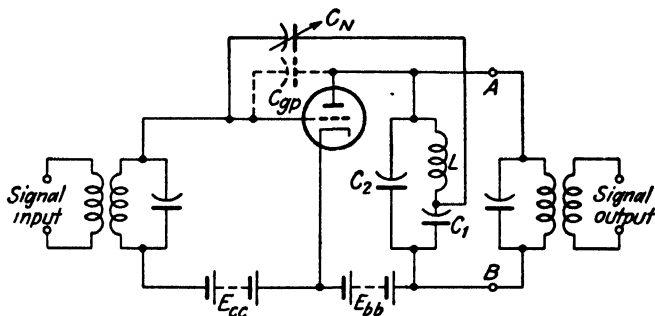



FIG. 195.—A method of neutralization in which the neutralizing voltage is obtained from the L-C_1-C_2 combination. The use of this separate circuit is more satisfactory than attempting to obtain the neutralizing voltage from the tuned output circuit.

This is accomplished by the circuit of Fig. 195. The circuit L-C_1 is adjusted so that inductive reactance predominates, and the signal current through L-C_1 lags the voltage across L-C_1 by 90°. This current will cause a voltage drop across condenser C_1 that will lag the current by 90°. This produces a voltage across C_1 that is 180° out of phase with the voltage across L-C_1. Now the voltage across L-C_1, which is the plate signal voltage of the tube, is the voltage that feeds current back through the interelectrode capacitance C_{gp}. Since the voltage drop across C_1 is 180° out of phase with the plate signal voltage, a variable neutralizing condenser C_N can be made to feed back a signal 180° out of phase. The two currents, the one fed back through C_{gp} and the one fed back through C_N, then produce opposite effects in the grid input circuit, and neutralization is accomplished. The equivalent reactance of the L-C_1 circuit is inductive, and condenser C_2 is used to tune the parallel combination to resonance so that the parallel combination offers high input impedance across the output circuit

A-B (from plate to cathode) and hence this parallel tuned circuit has little effect on the amplifier operation.

Sometimes the neutralizing voltage is obtained from the tuned load circuit. With such arrangements, complete neutralization may be difficult, and for most effective neutralization a separate tuned output circuit should be used across points *A-B*[1].

An alternate method of condenser neutralization is to feed back a signal from the same point that is fed back by the interelectrode capacitance C_{gp}, but to introduce the signal intentionally fed back at a point in the grid circuit so that cancellation and neutralization

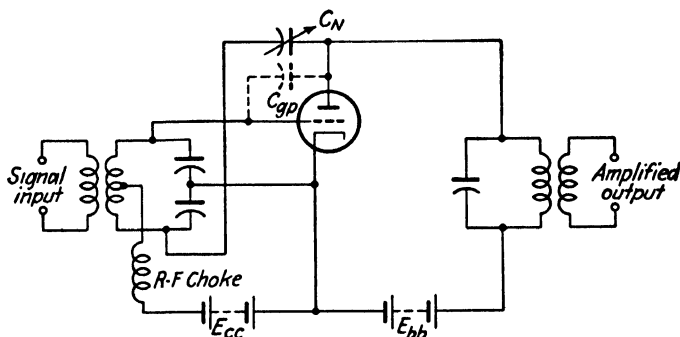



FIG. 196.—A method used sometimes for neutralizing a triode. This method often is not so satisfactory as that of Fig. 195. (See footnote.)

result. This arrangement is illustrated by Fig. 196. With the grid circuit balanced with respect to the cathode as indicated, the effect of the current fed back through the interelectrode capacitance C_{gp} is offset (at least partially) by the effect of the current fed back through the neutralizing condenser C_N. This condenser introduces its current at a point in the balanced circuit opposite to that of C_{gp}. As explained by Doherty in the footnote to the preceding paragraph, this method of neutralization is not too satisfactory; it is used, however.

Neutralization of Push-pull Amplifiers.—Radio-frequency power tubes often are operated in push-pull. This is particularly desirable where large power outputs are required. Such circuits are neutralized easily, because the voltages on the two plates are 180°

[1] See W. H. Doherty, Neutralization of Radio-frequency Power Amplifiers, *Pick-Ups*, December, 1939. This periodical was published formerly by Western Electric Company.

out of phase in the balanced push-pull circuit. This is because the plate current of one tube is rising when that of the other tube is falling. Thus signal voltages that are 180° out of phase are available without special arrangements, and the tubes can be effectively neutralized by the two condensers C_N of Fig. 197.

Radio-frequency Power Amplifiers Using Pentodes.—As has been mentioned several times, power amplifiers of small and intermediate sizes use power-output pentodes and beam-power tubes whose operating characteristics are similar to pentodes. Tetrodes

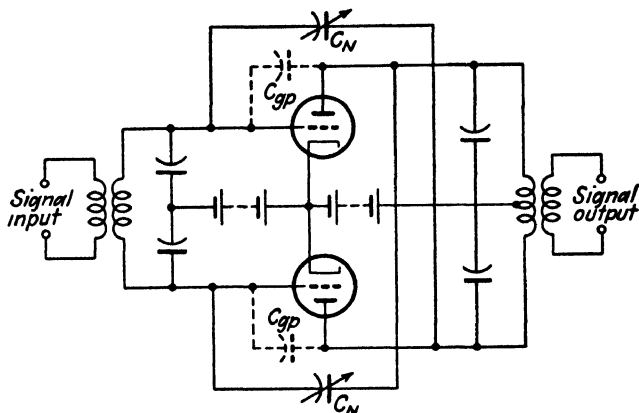



Fig. 197.—Method of neutralizing a push-pull radio-frequency power amplifier.

are used sometimes. Since these tubes operate in class B, or class C, and at radio frequencies, tuned parallel circuits are placed in the control-grid and plate-circuit leads.

Among the advantages of pentodes and beam-power tubes in radio-frequency power amplifiers are the following: (*a*) If the circuits are properly arranged and shielded (if necessary), then neutralization is not required at the radio frequencies at which these tubes ordinarily are operated. This has many advantages; neutralization is bothersome, particularly where the frequency of operation is to be changed frequently. (*b*) The grid-driving power that must be supplied by the preceding stage to a power amplifier using pentodes and beam-power tubes is less than that which would be necessary if triodes were used.

Controlled Feedback.—In this neutralization of triodes, two types of feedback are involved, uncontrolled feedback and controlled feedback. The signal fed back through the grid-plate

interelectrode capacitance is undesired, and it is not subject to control, except by the selection of a tube with low capacitance and by the nature of the external circuits involved. The neutralizing signal fed back to prevent oscillations is controlled by the size of the neutralizing condenser and by the points across which it is connected.

Feedback is of two types: **positive feedback** that adds to the applied signal voltage, *increases* the net signal voltage between grid and cathode, and tends to cause oscillations, and **negative feedback** that subtracts from the applied signal voltage and *decreases* the net signal voltage between grid and cathode. Positive feedback often is called **regeneration,** and negative feedback is called **degeneration.**

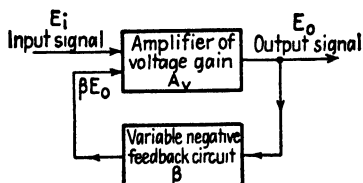



FIG. 198.—Block diagram for studying controlled feedback.

Because negative feedback reduces the actual net signal voltage that appears between grid and cathode, negative feedback reduces the output voltage and the effective amplification of an amplifier. Operating an amplifier in this way may seem absurd, but if an amplifier is designed with more output than is desired, and if the excess gain is then reduced with negative feedback, the distortion and noise level are lower than with ordinary amplifiers, and the gain is made quite independent of supply-voltage fluctuations.

The principle of negative feedback is illustrated in Fig. 198. The input signal voltage E_i is impressed on the amplifier of voltage gain A_v. Without feedback the voltage output will be A_vE_i. With negative feedback the gain will be reduced to E_O. The controlled negative feed-back circuit impresses a certain portion βE_O of this output voltage back to the input terminals so that the net signal voltage that appears between cathode and grid is reduced to $E_i - \beta E_O$. With negative feedback, the magnitude of the amplified output voltage is $E_O = (E_i - \beta E_O)A_v$. The voltage gain with negative feedback is the ratio of the output to input voltage, or

$$\text{Voltage gain with negative feedback} = \frac{E_O}{E_i} = \frac{A_v}{1 + \beta A_v}. \qquad (111)$$

For positive feedback the sign preceding β would be negative, thus

giving an increase in the voltage gain. Such an amplifier tends to be unstable and is likely to oscillate.

Although the preceding discussion largely applies to voltage amplifiers, negative feedback also is applied to power amplifiers. In such amplifiers, nonlinear distortion may result because the tubes often are worked to the limit of their capacities. Negative feedback reduces nonlinear distortion of the voltage output as follows:

$$\begin{matrix}\text{Distortion with} \\ \text{negative feedback}\end{matrix} = \frac{d}{1 + \beta A_v}, \qquad (112)$$

where d is the distortion without feedback. Methods of controlled feedback now will be discussed.

Voltage Feedback.—The circuit of Fig. 199 is arranged for *negative voltage* feedback. Assume for the moment that no signal is fed back. The impressed input signal voltage E_i will be amplified, producing an output voltage E_O across the load resistor R_L. This amplified output voltage E_O (which is the same as the voltage from plate to cathode) is 180° out of phase with the input signal voltage E_i (page 274). If condenser C is so large that its reactance is negligible, then the signal current that will be forced through R_1 and R_2 will be in phase with the amplified output voltage E_O. At a given instant, the relative polarities of the impressed voltage E_i and the voltage drop across R_2 can be found from the plus and minus signs. These indicate that the signal which is fed back and which exists as a voltage across R_2 will *subtract* from the applied signal voltage E_i. The feed-back factor β for Eqs. (111) and (112) is the ratio of the voltage E_2 across R_2 to the voltage output, or

FIG. 199.—A circuit arranged for negative voltage feedback. The signal fed back is the voltage drop across resistor R_2.

$$\beta = \frac{E_2}{E_O} = \frac{R_2}{R_1 + R_2}. \qquad (113)$$

Current Feedback.—The circuit of Fig. 200 is arranged for *negative current* feedback. Because the grid is negative, negligible current is forced through the grid circuit and cathode resistor R_c by the alternating signal voltage E_i applied between grid and cathode. As shown on page 274, the plate current I_p is *in phase*

with the grid voltage. Thus at the instant that the alternating grid signal voltage has the polarity indicated by the plus and minus signs, the instantaneous voltage drop across the cathode resistor R_c is as indicated. As these signs show, the voltage from cathode to grid will be the difference between the impressed signal E_i and the voltage drop across R_c. The two resistors R_L and R_c are in series, and hence the feed-back factor for Eqs. (111) and (112) is

$$\beta = \frac{R_c}{R_L}. \tag{114}$$

It will be recognized that this is the type of feedback that will occur if the cathode condenser is omitted from the self-biasing arrangement discussed on page 285. Such feedback exists to a limited extent at low frequencies if the cathode condenser is too small.

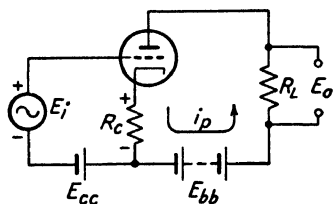



FIG. 200.—Circuit arranged for negative current feedback. The signal fed back is the voltage drop across resistor R_c.

The Cathode Follower.—If the load resistor R_L of Fig. 200 is omitted, the basic circuit of the cathode follower results. This is sometimes referred to as an amplifier, but it is not an amplifier in the strict sense of the word because the output voltage is *less* than the input voltage, which means that the gain is *less than unity.* The output terminals (not shown in Fig. 200) are across the cathode resistor R_c. The entire output voltage is fed back in series with the input voltage E_i. The circuit is, therefore, that of a negative feed-back voltage amplifier with a feed-back factor $\beta = 1.0$.

If the usual equivalent generator circuit is drawn for Fig. 200, the equation for the voltage "gain" can be written. Thus, from Eq. (79), page 273, the voltage across R_c can be found, and this, subtracted from the input voltage, gives the actual voltage impressed between grid and cathode for amplification. The use of this voltage as the actual signal voltage to be amplified gives an expression for the voltage drop across R_c. By solving for the ratio of the output voltage to input voltage the voltage gain is given as

$$A_v = \frac{E_o}{E_i} = \frac{\mu R_c}{r_p + R_c(1 + \mu)}. \tag{115}$$

This equation shows that the voltage gain is less than unity.

Cathode followers have the following important characteristics: (a) The grid-input capacitance is low in comparison with conventional amplifier stages, and the input impedance is a high value of essentially pure resistance. Because of this, cathode followers often are used between stages in wide-band video amplifiers so that the low capacitance improves the high-frequency response. (b) At frequencies below several million cycles the internal impedance between the output terminals essentially is a low value of pure resistance. Thus, with high-input impedance and low-output impedance the cathode follower is in a sense an impedance transformer, and can be used for this purpose. (c) The output voltage is in phase with the input voltage, and this sometimes is of importance.

It is of interest to point out that the plate of a cathode follower is sometimes grounded. It is also of interest to note that the cathode circuit of conventional amplifiers usually is grounded. Certain radio-frequency amplifiers for use at very high frequencies have grounded grids. Thus, in radio, tubes may be found operating with grounded cathodes, plates, or grids.

SUMMARY

Power amplifiers furnish the electric power that drives a loudspeaker, a radio-transmitting antenna, or some other electrical device. Because these amplifiers are power-output devices, they follow the theories of maximum power output and impedance matching. The equivalent generator circuit, so useful with voltage amplification, also applies to power tubes. The equation for power output is

$$P = \left(\frac{\mu E_g}{r_p + R_L}\right)^2 R_L.$$

Maximum power flows from a power tube to a load, when the load resistance equals the plate resistance. In radio apparatus distortion also must be considered, and maximum undistorted power output is obtained from triodes when the load resistance is about twice the plate resistance. For pentodes and beam-power tubes the load resistance is about one-tenth the plate resistance. Power-output transformers are used between the tube and load to match the impedance and obtain maximum undistorted power output.

Class A amplifiers operate with the grid biased so that operation is along the essentially straight part of the dynamic curve. The plate current closely resembles the grid signal voltage. Class B amplifiers operate with the grid biased to cutoff. The plate current is distorted badly, and is the same shape as the output of a half-wave rectifier. Class C amplifiers operate with the grid biased considerably beyond cutoff, and the plate current is even more distorted.

In the design of power amplifiers, dynamic instead of static curves are used. This is because when a tube amplifies, a voltage drop occurs across the load resistance and the plate voltage does not remain constant as when the static curves are taken. The dynamic curves are very convenient for showing how the grid signal voltage causes plate-current changes, but a family of constant grid-voltage curves and a load line are more convenient for calculations. For a single triode the power output is

$$P = \frac{(I_{\max} - I_{\min})\,(E_{\max} - E_{\min})}{8}.$$

For two tubes in push-pull the power output is

$$P = \frac{I_{\max}\,E_b}{5}.$$

The percentage of harmonic distortion and the value of load resistance can be determined from the equations given in the text.

The grid bias is very important in preventing distortion in class A operation. If the bias is too great, the grid may be driven past plate-current cutoff, and the peak of the negative half cycle of plate-signal current may be "chopped" off. If the bias is not sufficiently great, the grid may draw current on the positive half cycle, and if the source of signal cannot maintain the voltage, the positive half cycle may be "rounded" off. If the bias is correct, but the magnitude of the signal voltage is too great, then both peaks may be distorted.

Often in selecting a tube for a power amplifier, as small a tube as can be used is selected for reasons of high efficiency and economy. In audio amplifiers a single power-output tube sometimes is used, but often the tubes are operated in push-pull because with given tubes a greater output with the same distortion can be obtained. Tubes are operated in class A and class B for audio power amplification. If grid current flows, the designation A2 or B2 is employed. Both triodes and pentodes (and beam-power tubes) are used at audio frequencies. The plate-circuit efficiency is higher in class B than class A.

Radio-frequency power amplifiers are used for driving radio-transmitting antennas and for industrial heating purposes. Triodes, tetrodes, pentodes, and beam-power tubes all are used for radio-frequency power amplifiers with low and intermediate power outputs, but triodes are used for large power amplifiers.

Radio-frequency power amplifiers are usually operated in class B and class C, because large amounts of power often are handled and high efficiency is important. For amplifying a single frequency, such as an unmodulated carrier, class C operation is used, but for amplifying a band of frequencies, such as a modulated carrier, class B is used because of less distortion.

If the output circuit of a class C amplifier is tuned to a harmonic rather than to the fundamental, then the device becomes a frequency multiplier. Such circuits often are used. For best results the angle of flow should be

$$\theta_b = \frac{180^\circ}{n}.$$

Vacuum-tube voltmeters are very useful in studying the operation of radio-frequency power amplifiers and for other purposes. A peak voltmeter in the

grid circuit will measure the maximum positive value to which the grid voltage rises, and a trough voltmeter in the plate circuit will measure the minimum positive value to which the plate voltage rises From these measurements it can be determined if the tube is operating according to the selected T point. Operation at the selected Q point is determined by measurements with direct-current voltmeters The power output of a radio-frequency amplifier can be determined by measuring the signal-current input to a known resistance, or from the signal-voltage drop across a known resistance

Triodes used in radio-frequency amplifiers must be neutralized; otherwise feedback from the plate to the grid through the interelectrode capacitance may cause oscillations Neutralization is achieved by intentionally feeding back a signal such that the effect of the feedback through the grid-plate capacitance is canceled

Controlled feedback often is used across portions of amplifiers This may be positive feedback, which increases the net signal between grid and cathode and which may cause regeneration and oscillations, or it may be negative feedback, which decreases the net signal between grid and cathode and produces degeneration Controlled negative feedback reduces distortion and stabilizes an amplifier, making the effective gain quite independent of variations of supply potentials The voltage gain and distortion with negative feedback are given by the equations

$$\text{Voltage gain with negative feedback} = \frac{A_v}{1 + \beta A_v} \qquad \text{distortion with negative feedback} = \frac{d}{1 + \beta A_v}.$$

Controlled feedback may be voltage feedback or current feedback The cathode follower, a circuit used in television video amplifiers, is a feed-back circuit Its input impedance essentially is pure resistance of high value, and its output essentially is pure resistance of low value The gain of the cathode follower is less than unity, and is given by the equation

$$A_v = \frac{\mu R_c}{r_p + R_c(1 + \mu)}.$$

REVIEW QUESTIONS

1. In what important respects does a power amplifier differ from a voltage amplifier?
2. Is a class A1 amplifier a power amplifier? Discuss fully.
3. With triode power amplifiers, why is the load resistance about twice the plate resistance? With pentodes, why is the load resistance about one-tenth the plate resistance?
4. Why are power-output transformers used with power tubes?
5. Define Class A, B, and C operations. What is the meaning of a 1 or 2 following the letter?
6. What class of operation is used with voltage amplifiers? Why?
7. What is meant by angle of flow?
8. What is the relationship between a dynamic curve and a load line?
9. With a class A amplifier, the peak of one half cycle is "rounded" off. What is the probable cause, and how can it be remedied? What is the probable cause and what is the remedy if one half cycle is "chopped" off?

10. When the signal voltage is applied to a class A power amplifier, the direct-current milliammeter in the plate circuit reads a higher current than before the application of the signal. What does this indicate?

11. What are the advantages of pentodes over triodes in audio-frequency power amplifiers?

12. For audio-frequency power amplification, explain why two tubes in class B always are used.

13. Why are tuned input and output circuits used in radio-frequency amplifiers, but not in audio amplifiers for handling speech or music?

14. Explain why the load circuit of Fig. 188*a* may be represented by a resistance R'.

15. What is meant by the Q point? How would measurements be made to determine if the tube were operating in accordance with the selected Q point?

16. Repeat Question 15 for the T point.

17. For what main purpose is a class C amplifier used in radio?

18. What class of operation would be used for the final power amplifier of a radio-broadcast transmitter?

19. If the output of a transmitter were 50 watts, what type tube probably would be used? Why? Repeat if the output were 1.0 kilowatt.

20. Explain how a frequency multiplier operates.

21. Explain how the slide-back vacuum-tube voltmeter operates.

22. What is the function of the condenser across the milliammeter and battery of Fig. 191? Would the circuit operate satisfactorily if it were connected directly across the milliammeter?

23. How could the milliammeter of Question 10 be made to measure only the *change* in plate current?

24. A small radio-frequency power amplifier is supplying power to a load. How can the power delivered to the load be measured?

25. If the resistance of an unknown load is measured with a radio-frequency bridge, why may not the measured resistance values hold for computing the power?

26. What is meant by neutralization? What types of tubes usually need no neutralization, and why?

27. Why is coil neutralization completely satisfactory only for transmitters of fixed frequency? Does this limitation hold for condenser neutralization?

28. Would coil neutralization or condenser neutralization be used on a beam-power tube in a radio-frequency power amplifier operating at 1.75 megacycles fixed frequency? Explain your answer.

29. What is the basic principle and advantages of controlled negative feedback?

30. What is the cathode follower; what are its important characteristics, and what is one important application?

PROBLEMS

1. A power-output tube in a class A1 audio-frequency amplifier operates at a grid bias of -63 volts. What is the effective value of the grid signal voltage that will drive the tube to maximum output?

2. If the plate resistance of the tube of Prob. 1 is 1900 ohms, and the

amplification factor is 3.8, about what will be the maximum possible power output? The maximum undistorted power output?

3. If the tube just considered operates at a direct-plate voltage of 350 volts and direct-plate current of 45 milliamperes, what will be the approximate plate-circuit efficiency when the tube is delivering approximately maximum undistorted power output?

4. A type 2A3 triode is to be operated as an audio-frequency class A1 power amplifier with 250 volts on the plate, and with maximum undistorted power output. Obtain the static curves experimentally, or from a tube manual, and determine (*a*) the grid bias to be used; (*b*) the value of the cathode resistor; (*c*) the values of I_{max}, I_{min}, E_{max}, E_{min}, and I_b with 2500 ohms in the plate circuit; (*d*) the signal power output; (*e*) the direct power input; (*f*) the plate-circuit efficiency; (*g*) and the percentage distortion.

5. If two 2A3 tubes are used in a push-pull power amplifier operated in class A1, and are to be operated with 250 volts on the plate, what maximum undistorted power output can be obtained, and what plate-to-plate load resistance should be used? If the amplifier is to drive a 15-ohm loudspeaker, what should be the ratio of the output transformer?

6. Repeat the calculations starting on page 344 for a type 809 tube operated with 1000 volts on the plate. All other conditions are as specified in the illustrative problem. Also calculate the angles of flow for grid and plate currents.

7. A class C tube delivers 60 watts as a power amplifier. What should be the angle of flow, and what will be the approximate output when this tube is used as a frequency tripler?

8. The grid-plate capacitance of a large power triode is 6.4 micromicrofarads. What size coil and condenser are recommended for coil-neutralizing this tube at 2.1 megacycles? Consult a radio-supply catalogue for a standard capacitor. The plate voltage is about 2000 volts *direct voltage*, and the direct grid voltage is about -500 volts.

9. The amplifier of Fig. 199 has a gain without feedback of 12.5. What will be the approximate gain with feedback if $R_1 = 0.25$ megohms, and $R_2 = 10{,}000$ ohms?

10. If Eq. (115) is correct, prove numerically that the voltage gain of a cathode follower is less than unity.

CHAPTER X

OSCILLATORS

In both audio-frequency and radio-frequency apparatus, sources of alternating-current electric energy often are required. Such sources usually are vacuum-tube oscillators. For audio purposes, and for the usual radio frequencies, ordinary vacuum tubes are used, but for the ultrahigh frequencies special tubes have been developed.

It has been explained in preceding discussions that a small signal voltage impressed on the control grid of a tube is amplified and appears as a larger voltage, equal to μE_g, in the plate circuit. Also, it has been explained that the control-grid circuit draws a negligible amount of power, or but little power, compared with the power that may be delivered by the plate circuit. Thus, if some of the output signal voltage is fed back into the control-grid circuit in the proper manner, a vacuum tube will oscillate. Furthermore, since but a small amount of power need be fed back, a relatively large percentage of the power available may be delivered to some external circuit.

The subject of feedback was considered in the preceding chapter. For a tube to oscillate, the feedback must be such that the oscillations will build up. It was explained that oscillations will occur with triodes that are not neutralized. In most oscillators, however, a separate feed-back circuit is used, so that the oscillations can be controlled. In a sense, the vacuum-tube oscillator is a converter because it changes the direct-current power impressed on the electrodes into alternating-current power.

Types of Oscillators.—An examination of radio literature will reveal that a large number of apparently different types of oscillators have been developed. Certain of these oscillators when examined closely will be found to differ but little from one another. In this chapter oscillators will be considered from as fundamental a standpoint as appears possible, and of course, it will be possible to consider only the important basic types.

Vacuum-tube oscillators may employ triodes, tetrodes, pentodes,

and beam-power tubes. These three types are used in low-power and intermediate-power oscillators. Triodes are used in large oscillators delivering more than, perhaps, a thousand watts.

Circuits possessing *negative resistance* may oscillate, and this principle sometimes is used to explain oscillations in vacuum-tube circuits. Positive resistance is resistance of the ordinary type that *absorbs* power; negative resistance is of the type that *delivers* power under certain conditions. For example, as indicated in Fig. 104, page 189, between the plate voltages of 10 and 50 volts an *increase* in voltage causes a *decrease* in plate current. This is opposite to the usual resistance effect where an increase in the voltage across a resistor causes an increase in the current through the resistor. The tetrode of Fig. 104 therefore possesses negative resistance over the region from 10 to 50 volts, and because of this tendency to return power to the circuit in which it is connected the tetrode can be made to oscillate in a circuit called a **dynatron oscillator.** The **transitron oscillator** also uses a tetrode and operates on the negative-resistance principle. Because these devices are relatively unimportant they will not be discussed further.

The oscillators that largely will be considered are of the feedback type in which a separate circuit is used to feed back signal from the plate to the grid circuit. There will be exceptions to this statement, however, because in many oscillators the signal feedback is through the interelectrode grid-plate capacitance instead of through an external circuit. Other exceptions will be for ultrahigh and superhigh frequencies.

Oscillators may be classified on the basis of the frequency that they generate. A convenient classification is as follows:

Audio-frequency oscillators—20 to 25,000 cycles per second.

Radio-frequency oscillators—25,000 to 30 billion cycles per second. This classification is entirely arbitrary. It is based on the fact that frequencies from about 20 to 25,000 cycles are in the audible range, and those above about 25,000 cyles are inaudible to the average human ear. The upper limit of 30 billion cycles is of course not the upper limit at all, but merely the approximate upper frequencies in which present extensive investigations are under way. This limit rapidly is being pushed upward.

For further classification of radio-frequency oscillators the ranges specified by the Bureau of Standards[1] will be used.

[1] *Electrical Engineering*, Vol. 65, No. 5, p. 238, May, 1946.

Classification	Abbreviations	Frequency (megacycles)	Wavelength (centimeters)
High frequency	H.F.	3—30	10,000—1000
Very high frequency.......	V.H.F.	30—300	1000—100
Ultrahigh frequency.......	U.H.F.	300—3000	100—10
Superhigh frequency.......	S.H.F.	3000—30,000	10—1

Audio-frequency Oscillators.—These are used extensively as sources of electric energy of adjustable frequency for routine tests on amplifiers and other speech-input equipment, and for experimentation. Three basic types of oscillators are in wide use: (*a*) those which employ tuned resonant circuits of inductance and

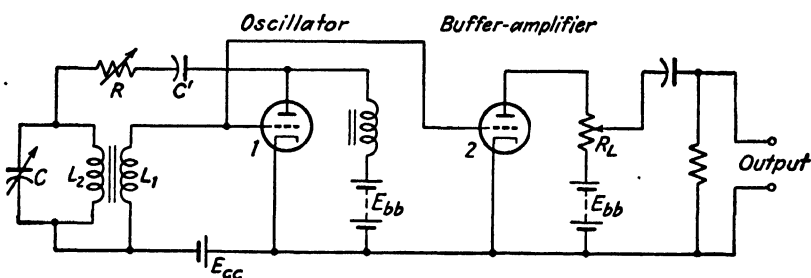



FIG. 201.—Circuit of an audio-frequency oscillator. The output terminals may be connected to additional voltage-amplifying stages or to a power-output tube. An audio-frequency transformer may be substituted for R_L.

capacitance in parallel, and which are often called **L-C oscillators**; (*b*) those which employ a circuit of resistance and capacitance to determine the frequency of oscillation, and which are known as **R-C oscillators,** and (*c*) the so-called **beat-frequency oscillators.** The first two oscillators will be discussed in the pages immediately following. The beat-frequency oscillator will be discussed on page 425 because it will be more easily understood after the subjects of modulation and demodulation have been studied.

Inductance-capacitance (*L-C*) *Oscillators.*—Many circuit possibilities exist. The discussion will be confined to the oscillator of Fig. 201 which is the basic circuit of an audio-frequency oscillator much used for test and experimentation.

Tube 1 is the oscillator. An audio-frequency choke coil is placed in the plate circuit. This takes the place of the usual resistor used as the load in the plate circuit of an amplifier (remember that an oscillator is an amplifier with feedback). The inductor is used

instead of a resistor so that there will be little opposition to the flow of direct plate current. There will be an alternating signal voltage across the inductor, similar to the signal-voltage drop across the load resistor of an amplifier. For this to occur, the inductor should offer considerable inductive reactance at the lowest frequency at which the oscillator is to operate. A coil having an inductance of about 15 henrys is satisfactory for most purposes (remember this must be the incremental inductance at the value of plate current flowing).

The circuit composed of C', R, C, and L_2 is connected across the inductor in the plate of tube 1. Thus the voltage across this inductor forces a current through C'-R in series with C-L_2 in parallel. Condenser C' is a blocking condenser to isolate the plate and grid circuits from the direct voltage. It should have little reactance at the frequency of oscillation compared with the resistance of resistor R. It may be a good paper capacitor of about 1 microfarad capacitance and capable of withstanding several times the plate voltage. The variable rheostat R controls the amount of feedback. The maximum value of R is theoretically

$$R = R_L (\mu - 1) - r_p, \tag{116}$$

where R is in ohms, when R_L is the equivalent resistance in ohms at resonance of the parallel C-L_2 circuit (page 84), μ is the amplification factor of the tube, and r_p is the plate resistance in ohms. For a typical amplifier, R as given by Eq. (116) would be of the order of about 100,000 ohms. This is a theoretical value and usually needs adjustment.

The tuned circuit C-L_2 largely determines the frequency of oscillation. The losses in the coils and the iron core should be low so that the tuned circuit will have a high Q (page 49) and be sharply tuned to the frequency desired. Using typical circuit elements, the effective resistance of the tuned parallel circuit C-L_2 is of the order of 10,000 ohms or more.

In accordance with the theory of the parallel circuit, the small signal current fed back through C'-R will cause a voltage drop across the parallel circuit C-L_2, and a rather large current of the resonant frequency will flow in L_2 and will induce a voltage $E = \omega M I_p$ (page 86) in the coil L_1 connected between the grid and cathode. This voltage will be amplified, and will feed back a signal through C'-R, etc., and the circuit will oscillate. Of course the

phrase relations must be correct, and if the circuit does not oscillate, one of the first things to do is to reverse the connections of the secondary L_1. The coils L_2 and L_1 may be the windings of a one-to-one-ratio audio-frequency transformer. A transformer used to isolate one audio-frequency transmission line from another (that is, designed to work from about 500 ohms into 500 ohms) is satisfactory in the 1000-cycle region, but of course other transformers will operate with excellent results. In fact, if a wide range of frequencies is to be covered, several transformers and a bank of capacitors is necessary.

The oscillator tube 1 should be of the type that has an amplification factor of about $\mu = 10$ and a plate resistance of about $r_p = 10{,}000$ ohms, although of course other tubes will work with satisfaction. Tube 1 is biased to operate as a class A amplifier, but in actual operation the oscillations build up until the grid is driven slightly positive on each positive half cycle, and grid current flows. Thus operation is in class A2. In fact, it is this action that limits the magnitude of the oscillations. When the tube is energized, a transient condition exists, and the oscillations that build up are at the frequency largely determined by the frequency of resonance of the parallel C-L_2 circuit. Of course these oscillations cannot build up indefinitely. When grid current flows, the grid circuit draws power, and the grid circuit accordingly becomes resistive. This couples resistance back through L_1 into L_2 in accordance with coupled-circuit theory (page 101), and for this reason the primary current in L_2 that induces the oscillating voltage in L_1 ceases to increase and a steady-state condition is reached. The frequency of oscillation is approximately

$$f = \frac{1}{2\pi\sqrt{L_2C}}, \tag{117}$$

where f is in cycles per second, when L_2 is in henrys and C is in farads. This equation applies only approximately to parallel circuits (page 81); furthermore, a complete analysis shows that such factors as the tube coefficients enter into determining the frequency. Also, the magnitudes of these coefficients are determined by the electrode potentials. Nevertheless, Eq. (117) is satisfactory because such an oscillator is calibrated, and settings of C, L_2, and R are made for each output frequency desired.

It will be noted that the grids of tubes 1 and 2 are connected.

This means that whatever alternating signal voltage is on the grid of tube 1 also appears between the grid and cathode of tube 2. Thus if tube 1 (and tube 2, as indicated in Fig. 201) is biased to -3 volts, the peak value of the alternating signal voltage will be about 3 volts, assuming that the grid is driven slightly positive when the oscillations become stable.

Tube 2 usually is a voltage amplifier of the same type as tube 1. Tube 2 may be resistance coupled to the output terminals, or it may be transformer coupled. It may drive another voltage amplifier, or it may drive a power amplifier. If the signal voltage between the grid and cathode is estimated as explained in the preceding paragraph, then the voltage, or power output, of tube 2 can be determined as for any audio amplifier, as explained in the preceding chapters. Tube 2 also serves as a buffer, that is, as a tube to *isolate* the oscillating tube from the output terminals. With the circuit of Fig. 201, the usual types of loads connected to the output terminals, and to which power is delivered, have negligible effect on the frequency of oscillation. But if power were drawn directly from the oscillating tube, then the frequency would vary, unless the amount of power taken were very small. Although separate plate-supply batteries are shown in Fig. 201, the two tubes can be operated from the same source.

Resistance-capacitance (R-C) Oscillators.—This device essentially is a resistance-coupled amplifier with controlled feedback applied in a manner such that sustained oscillations result. It is possible to use only one tube, but usually two tubes are used in the circuit of Fig. 202. Two stages produce a 360° phase shift between input and output voltage (page 274).

In studying Fig. 202, assume for the moment that the device is an amplifier in which an input signal voltage E_i produces an amplified output signal voltage E_O. Now E_i need not be impressed on the device from an external source, but may be obtained from the circuit itself as in any oscillator. This is accomplished by the feed-back circuit R_1-C_1-R_2-C_2. The voltage E_f fed back into the control-grid circuit of the first tube is equal in magnitude and phase to E_i, the voltage necessary to produce the amplified output voltage E_O. For proper adjustment of the feed-back circuit, the frequency of oscillation is

$$f = \frac{1}{2\pi\sqrt{R_1R_2C_1C_2}}, \qquad (118)$$

where f is in cycles per second, when the resistances are in ohms and the capacitances in farads.

The lamp R_3 acts like a ballast tube (page 251), and in combination with resistor R_4 it provides a stabilizing circuit that controls the amplitude of oscillations and makes the oscillator characteristics largely independent of supply voltages and tube characteristics. It will be noted that the ballast lamp R_3 is between the control grid and cathode of the first tube and that it causes negative feedback in accordance with the explanation on page 364.

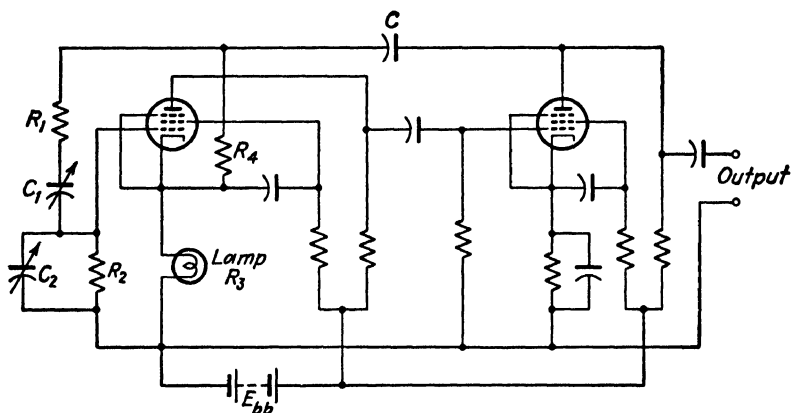



Fig. 202.—Circuit for a common resistance-capacitance oscillator.

Radio-frequency Oscillators.—Of course, if the circuit constants are properly selected, the circuits of Figs. 201 and 202 will oscillate in the radio-frequency range. These circuits seldom are used primarily for radio-frequency purposes, however, although the oscillator of Fig. 202 is used up to 200,000 cycles and above.

Radio-frequency oscillators usually operate with the tubes in class C because of the greater efficiency and because at radio frequencies the harmonics resulting from class C operation are suppressed readily by simple tuned circuits (page 80). The general classification of radio-frequency oscillators covers a very wide range. No single circuit is satisfactory for all frequencies. In recent years the upper frequency limit of radio has been pushed back farther and farther, until now wavelengths of a fraction of a centimeter and frequencies of thousands of megacycles are produced.

Basic Class C Radio-frequency Oscillator.—This is shown in Fig. 203. When the circuit is not oscillating, the grid bias is zero. When the circuit is energized by connecting the electrode potentials, an alternating current, the frequency of which is determined to a great extent by the tuned circuit L_2-C, flows in the plate circuit. In accordance with the explanation on page 80, a relatively large current at this resonant frequency will flow through coil L_2, and this will induce a voltage $E_g = \omega M I_2$ in coil L_1, connected between the grid and cathode of the tube. If the coil L_1 is correctly connected, this voltage E_g will be amplified, and a larger current I_2 will result. If it is not properly connected, the voltage fed back will be of the wrong phase relation and the oscillations will not build up. If a circuit such as Fig. 203 does not oscillate, the connections of coil L_1 should be reversed.

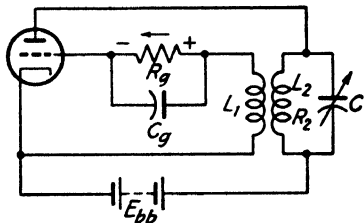



Fig. 203.—Basic circuit of the self-biased class C radio-frequency oscillator.

As was mentioned, when the circuit is not oscillating, the bias, or direct voltage, between the grid and cathode is zero. As soon as oscillations start, the grid is driven positive and grid current will flow on the positive half cycles as for a class C amplifier (page 343). This current will flow through the grid resistor R_c and will cause an $I_c R_g$ voltage drop, which has the polarity shown in Fig. 203. This direct voltage is between the grid and cathode and thus biases the tube and fixes the point of operation for class C. Capacitor C_g makes it possible for the alternating voltage induced into L_1 to appear between grid and cathode; otherwise much of the signal voltage would be lost in R_g. Capacitor C_g, shunting resistor R_g, forms in a sense a capacitor-input load across the grid circuit, which acts like a dioder ectifier (page 233). Because of this condenser-input action the presence of the condenser affects the direct current and voltage in the grid circuit.

When negligible current flows in the grid circuit the input impedance between grid and cathode of the tube of Fig. 203 is very high, and the grid circuit draws negligible power from the plate circuit. When the grid is driven positive by the voltage that is fed back by the coil, then the grid circuit draws power, and its

equivalent input resistance drops. This couples a value of resistance back into coil L_2 that combines with the actual resistance, producing the equivalent resistance R_2. This will reduce the magnitude of the alternating current that flows in coil L_2 and of the voltage induced in L_1. By this action the oscillations become stabilized. The frequency of oscillation is approximately

$$f = \frac{1}{2\pi\sqrt{L_2C}}, \tag{119}$$

where f is in cycles, when L_2 is in henrys and C is in farads. An alternating radio-frequency signal voltage can be obtained by coupling another coil with coil L_2, or by suitable connections to the tuned L_2-C parallel circuit. If power is drawn, this is equivalent to changing R_2 and this will affect the frequency to which the circuit is tuned (page 81). Such action is not predicted by Eq. (119), because this equation is approximate. In oscillators the desired frequency usually is obtained by slight adjustments of the tuning capacitor C.

The design of a class C radio-frequency oscillator follows the method outlined for the class C radio-frequency amplifier (page 343). The value of the grid resistor R_g can be determined from the value of grid bias desired and the direct grid current that flows. These are determined by the T and Q points selected. The effective Q of the parallel load circuit $Q = \omega L_2/R_2$ should be about 12 for stable operation and good efficiency. Note that R_2 includes the effect of the power drawn by the grid circuit and the connected load circuit (page 102).

The capacitance of the grid condenser C_g is not critical, but it should offer low reactance to the frequency of oscillation given by Eq. (119). Because such wide deviations exist among tubes it is not advisable to attempt to specify values of R_g and C_g; instead it is suggested that the values specified in circuits recommended by manufacturers in their tube manuals and the values given in handbooks be used.

Intermittent operation may result if the values of R_g or C_g are too great. During normal operation the oscillator reaches an equilibrium condition and the bias reaches a constant value. If, now, the load is altered and the power drawn is changed, equilibrium conditions are upset, and a new state of operation with a new grid bias must be established. If the grid resistor R_g and the

grid condenser C_g are too large, the condenser may not be able to lose its charge sufficiently fast (of course the direct-current voltage to which it is charged is I_cR_g, as explained early in this section). Thus the grid may be biased momentarily too far negative to satisfy the new condition of operation, and the oscillator may miss a few cycles until the direct grid potential falls to the proper value. Ordinarily, intermittent operation results only in erratic performance.

Blocking is a more serious phenomenon of oscillators, and may result in serious damage to tubes and to equipment, such as indicating instruments for measuring the electrode currents. During normal operation the grid goes positive and draws electrons from the cathode. These flow through the grid resistor R_g and produce the grid bias, as shown in Fig. 203. Now a given bias ($E_c = I_cR_g$) can be obtained from a small current and large resistor, or vice versa. When the grid is driven positive, secondary emission (page 182) from the grid occurs. Most of these liberated secondary electrons will be drawn over to the more positive plate and then back through R_g to the grid in a direction opposite to the normal grid current I_c. If R_g is too large so that I_c is small, and if operation is such that the grid is driven highly positive, this secondary current through R_g may exceed the normal current through R_g. As a result, the grid bias may be changed from negative to positive. This will permit the positive plate to draw a large current from the space-charge region, and this large current may cause the tube to overheat, may cause a liberation of gas from the electrodes, or may cause the electrodes to be warped or burned. The grids of modern tubes are treated to reduce secondary emission. Oscillators, when first energized for a given service, should be closely watched to ascertain if blocking will occur.

Simple oscillators similar to Fig. 203 are widely used in industry for heating purposes. Metal objects are heated by eddy currents induced by magnetic fields produced by such oscillators, which have frequencies of the order of several hundred thousand cycles. Nonconductors, such as wood, plastics, and food products, are heated by the alternating electric fields produced by such oscillators. The frequencies used ordinarily are from several million to about 50 million cycles.

Oscillator Circuits.—A large variety of these exist. In some oscillators the differences are in detail only. Thus the circuit of

Fig. 203 has the tuning in the plate circuit. It is sometimes known as a **tuned-plate oscillator.** If, on the other hand, the capacitor is omitted from the plate circuit and is placed in the grid circuit across coil L_1, then the oscillator becomes a **tuned-grid oscillator.** If condensers are connected across both L_2 and L_1 the circuit becomes that of a **tuned plate-tuned grid oscillator.** In radio, the feed-back coil of an oscillator sometimes is called a **tickler,** a term not so much used as formerly.

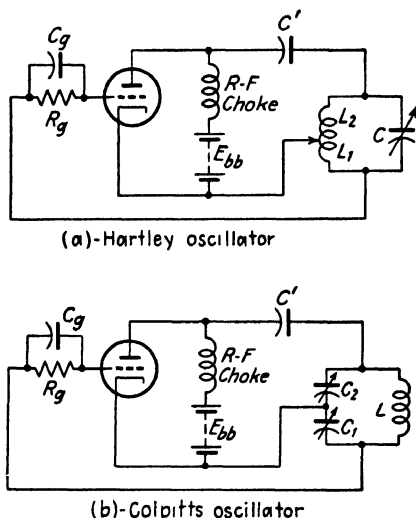



FIG. 204.—Typical radio-frequency oscillator circuits.

Hartley Oscillator.—The basic circuit arrangement is shown in Fig. 204*a*. The plate is "shunt fed" through the radio-frequency choke coil that keeps alternating current from flowing in the power supply connected at E_{bb}. The capacitor C' isolates the grid circuit from the plate supply E_{bb}. The tuned-*plate* circuit between plate and cathode is composed of two parallel arms. One is C and L_1 in *series*, and the other is L_2. Because L_1 is in series with C, the frequency of oscillator occurs *approximately* when the capacitive reactance of C minus the inductive reactance of L_1 equals the inductive reactance of L_2. The amount of feedback to the grid circuit is varied by changing the setting on coil L_1-L_2, which functions as an inductive voltage divider.

Colpitts Oscillator.—This is shown in Fig. 204*b*. Here the oscillating circuit between plate and cathode is composed of the series

arm L-C_1 in parallel with C_2. The frequency of oscillation will be approximately at a frequency such that the inductive reactance of L minus the capacitive reactance of C_1 equals the capacitive reactance of C_2. The amount of feedback to the grid circuit is varied by changing the settings of C_1 and C_2, which function as a capacitive voltage divider.

Crystal-controlled Oscillators.—The frequencies at which the oscillators of the preceding pages operate are controlled by tuned circuits of inductance and capacitance. In the design and construction of such oscillators precautions are taken to ensure that the frequency remains as constant as possible, but even so, the frequency stability is not sufficiently great for much modern radio apparatus. For instance, the basic carrier frequency of many radio transmitters is controlled by crystals of quartz.

This method of control is based on the **piezoelectric effect** present in quartz crystals. Thus, suppose that a piece of quartz is properly cut and ground into a thin slab *about* 1/16 inch thick and the size of a postage stamp. Furthermore, suppose that this slab of crystal is placed between two flat metal electrodes in a suitable holder and that an alternating voltage of variable frequency is connected across the electrodes. If, now, the frequency is varied, and the input impedance to the crystal and its other characteristics are studied, it will be found that at some frequency the crystal vibrates mechanically and that its electrical characteristics are those of a *very sharply tuned* parallel resonant circuit (that is, to a circuit with a very high Q, much higher than can be obtained with coils and condensers). Since a tuned parallel circuit between grid and cathode will control the frequency of oscillation, a properly cut and ground quartz crystal suitably connected in the grid circuit also will control the frequency of oscillation.

Cuts Used.—Quartz as it occurs in nature sometimes is in the form of crystals having major dimensions of several inches. These "raw" crystals are examined by several optical means and their optical axes, determined by the ways in which the molecules are aligned, are found. There are several ways of cutting the crystals along these optical axes, depending on the final crystal characteristics desired. After being cut on suitable saws, the crystals are ground and etched to a thickness such that they have the exact operating frequency desired. For the higher frequencies, quartz

crystals must be very thin, so thin in fact, that they may not have the required mechanical strength. The upper limit of the basic frequency of oscillation is *about* 15 megacycles.[1]

The oscillations of a quartz crystal may become so violent that the crystal will shatter as did the crystal shown.

Temperature Effects.—These are determined largely by the type of cut used. The **frequency-temperature coefficient** is the number of cycles of frequency change per megacycle per degree centigrade temperature change. This coefficient may be positive (increase in resonant frequency with increase in temperature) or the coefficient may be negative (decrease in resonant frequency with decrease in temperature). Crystals are cut in special ways so that the frequency-temperature coefficient is very small. Sometimes crystals are mounted in automatically controlled electric ovens to stabilize their operation. If crystals are driven too hard by applying an excessive resonant voltage across them, they will heat. If the voltage is sufficient, the oscillations may become so violent that the crystals will shatter.

A quartz crystal in a typical mounting of type (*b*) (see text), in which both metal electrodes touch the corners of the crystal. The holder and crystal have been cut away to show the interior. (*Bliley Electric Co.*)

Crystal Mountings.—The mountings, or holders, in which crystals are placed are of two general types: (*a*) those holders in which the crystal is in direct contact with the lower metal electrode but which have an air space between the crystal and the upper metal electrode, and (*b*) those holders in which both electrodes are in contact with the crystal. The first method gives a small amount of frequency control by

[1] Direct crystal control, using a crystal that has a signal frequency output as high as 50 megacycles, has been announced. See G. M. Thurston, Direct Crystal Control at 50 MC, *Western Electric Oscillator*, July, 1946. For useful information on quartz crystal control at high frequencies, see also G. B. Sells, V.H.F. Crystal Oscillators, *QST*, November, 1947.

varying the air gap. Of course the operation of the crystal is affected to a slight extent by the holder, and a crystal should be calibrated in its holder for precise results.

Crystal Oscillators Using Triodes.—A typical circuit is shown in Fig. 205*a*, and the equivalent circuit in Fig. 205*b*. The radio-frequency choke prevents the flow of alternating current through resistor R_1. The oscillations build up until the grid is driven positive, and the direct current flowing through R_1 biases the tube.

(a)-Actual circuit

(b)-Equivalent circuit

FIG. 205.—Actual and equivalent circuits for one type of triode crystal oscillator. The *C-L* circuit should be designed to tune to the crystal frequency, a typical value for *C* being 100 micromicrofarads. The value of R_1 is from 2500 to 25,000 ohms depending on the grid current that flows for a particular tube. A radio-frequency milliammeter or a radio dial light often is connected in series at point *X*.

The signal voltage and power required to maintain the crystal oscillations are fed back from the plate to grid circuit through the interelectrode capacitance between the plate and grid (page 287), which, it will be recalled, is high for a triode. In the final adjustment the parallel plate circuit is tuned to a frequency slightly higher than the resonant frequency of the crystal so that the power required to drive the crystal will be fed back (page 290).

In adjusting the crystal oscillator, the plate milliammeter will indicate when oscillations start. If the circuit is not oscillating, and if the capacitance of the tuning condenser is varied, the plate current will drop at a critical value at which oscillations start, as shown by Fig. 206. For best results the oscillator is operated over the region *B-C*. If the oscillator is loaded and delivering power, the dip in the direct plate current will not be so great. In regard to the dip in plate current, it should be noted that the grid is without negative bias until oscillations occur. A small neon bulb held near the tuned plate circuit also will indicate the presence of oscillations. The coils and condensers used in such oscillators

ordinarily are of the open air-insulated type. A crystal oscillator usually operates into several class C amplifiers, which act as buffers (page 375) and as power amplifiers.

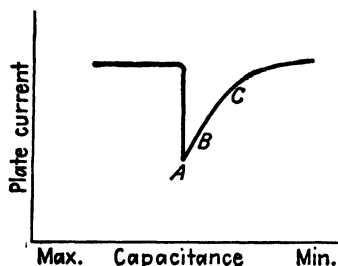



FIG. 206.—If the crystal of Fig. 205 is not oscillating and the capacitance of tuning condenser *C* is varied, the plate current suddenly will drop to point *A*, and oscillations will start. The circuit usually is then adjusted until operation is over the region *B-C*.

A radio-frequency voltage is fed back into the crystal circuit to cause the crystal to oscillate, and a radio-frequency current therefore flows in the crystal circuit. Safe currents are of the order of 100 milliamperes. Sometimes a flashlight bulb or radio-dial light is connected in the ground side of a crystal circuit at point *X* (Fig. 205*a*) to act as a fuse and to protect the crystal. The size of the light selected is such that it will burn out before the crystal is damaged. Also, the light serves as an indication of the magnitude of the oscillations which may ruin the crystal if excessive.

Crystal Oscillators Using Tetrodes and Pentodes.—These are used very extensively in radio equipment, a typical arrangement

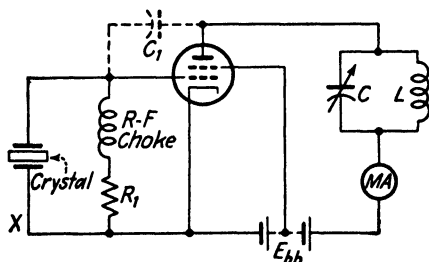



FIG. 207.—A crystal oscillator circuit using a tetrode. The value of the plate voltage is about 250 volts, and the screen grid is tapped on at about 100 volts. The value of R_1 is from 10,000 to 50,000 ohms. Of course a power supply with suitable dropping resistors can be used instead of batteries. In this event the lower end of the *C-L* circuit and the screen grid should be connected to the cathode through capacitors of the order of 0.001 microfarad. A radio-frequency milliammeter or a dial light may be placed at point *X* to indicate oscillations and to protect the crystal. Because of the very small interelectrode capacitance, it may be necessary to add a very small condenser as indicated by the dotted lines.

for a tetrode being shown in Fig. 207. Power-output tetrodes, pentodes, or beam-power tubes may be used. If the tube is a pentode having a suppressor grid, this grid is grounded. The

circuit requires little explanation; it is essentially a power amplifier with feedback through the (very small) interelectrode and wiring capacitance. Because this is so small, it may be necessary (especially with pentodes) to add condenser C_1, which should be very small, having a capacitance of but several micromicrofarads. A circuit such as Fig. 207 will give a greater power output than the circuit of Fig. 205 using comparable crystals and tubes.

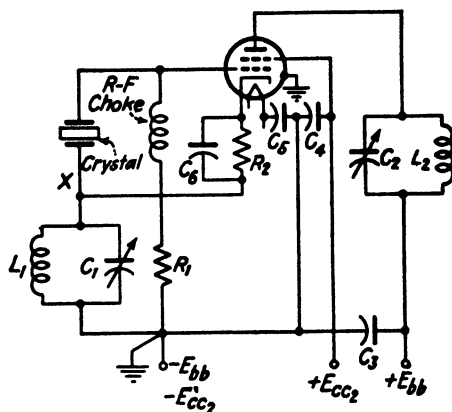



FIG. 208.—A crystal oscillator using a beam-power tube that acts both as an oscillator and buffer amplifier. Condensers C_3, C_4, C_5, and C_6 are not critical and are from 0.001 to 0.01 microfarad. Resistor R_1 is from 20,000 to 100,000 ohms, and R_2 is about 400 ohms. (*Circuit adapted from Radio Amateur's Handbook.*)

Electron-coupled Oscillators.—One important advantage of this circuit is that the screen grid is used as a shield between the oscillating circuit and the output circuit. This largely prevents changes in the power output from affecting the frequency of oscillation; hence, the tube provides its own "buffering." Since power tetrodes, pentodes, or beam-power tubes are used, the power output obtained with a given crystal is relatively high. A typical electron-coupled oscillator using a tetrode or beam-power tube is shown in Fig. 208. This is sometimes called a **tri-tet oscillator.**

The oscillating circuit consists of the cathode, control grid, screen grid, the crystal, and condenser C_1 and coil L_1. The screen grid is connected to a positive direct voltage as usual, but there is no alternating-current load (tuned circuit) in the screen-grid circuit, and because of condenser C_4 the screen grid is at zero radio-frequency voltage. The tuned load circuit L_1-C_1 is in series with the cathode, and signal feedback to the crystal occurs through the control grid-cathode interelectrode capacitance.

The control grid, screen grid, and cathode act much like any triode crystal-controlled oscillator. Thus the control-grid voltage variations cause corresponding alternating signal variations in the electron current flowing to the plate. A tuned load circuit L_2-C_2 is placed in the plate circuit, and these plate-current signal variations cause corresponding output voltage variations. In this way oscillations occur in the control-grid circuit, and alternating-current power is available in the plate circuit.

When the tube is operated at reasonably high plate voltages such that the plate current is but little affected by plate voltage (Fig. 104, page 189), then variations in the plate output circuit will not be reflected back into the control-grid oscillating circuit because of the shielding afforded by the screen grid. Thus both buffer action and relatively large power output are obtained from a single multielectrode tube.

Vacuum Tubes at Very High Frequencies.—Ordinary vacuum tubes designed for power output purposes and of the type used in radio-frequency power amplifiers and oscillators operate quite well up to about 15 megacycles. Above this frequency, the power that can be handled drops off quite rapidly. For example, a transmitting triode of standard construction with all leads brought out at the base will operate at full rating up to 15 megacycles, and at reduced ratings up to 85 megacycles. If minor changes are made, such as bringing the plate lead out of the top, etc., a transmitting triode of a standard type will give 100 per cent output at 60 megacycles, about 88 per cent output at 70 megacycles, and 50 per cent output at 120 megacycles. At very high frequencies (30 to 300 megacycles, page 372) the plate-circuit efficiency drops, the tube heats, and eventually it will cease to amplify or oscillate. The reasons for unsatisfactory operation at very high frequencies if more or less standard construction is used are many and varied. Important among these are the following:

Transit Time.—At the lower radio frequencies, the **transit time,** or time required for an electron to pass from cathode to plate, is of little importance because this time is negligibly small compared with the time required for one cycle of plate voltage to occur. In an amplifier or an oscillator an alternating voltage (as well as a direct voltage) exists between cathode and plate, and at very high radio frequencies this voltage is changing very rapidly. For example, at 100 megacycles, since $t = 1/f$, a complete cycle occurs

in 0.01×10^{-6} second. Thus, an electron (current) may start out to the plate with certain phase relations but arrive at the plate with incorrect phase relations because during the time of transit the alternating plate voltage has changed markedly. In a conventional tube the transit time is about 0.001×10^{-6} second. Among the effects of transit time are (*a*) an increase in the power drawn by the grid, (*b*) increase in temperature of the electrodes, particularly the plate, (*c*) decrease in plate-circuit efficiency, and (*d*) inability to operate as the frequency is increased to the upper limit. Transit time can be decreased by reducing the spacing between cathode and plate and by increasing the direct-plate voltage, which will speed-up the electrons.

FIG. 209.—Showing the interelectrode capacitances and lead inductances of a triode. These influence greatly high-frequency operation.

Internal Inductance and Capacitance.—The equivalent circuit of a triode is shown in Fig. 209. An examination of Eq. (117), which gives the approximate frequency of an oscillator, indicates that the inductance and capacitance must be *very* small for operation at very high frequencies. As Fig. 209 indicates, there is inductance in the leads and electrodes, and capacitance between the leads and electrodes, and if these were connected directly together at the tube base, these "residual" effects would determine an upper limit of oscillation. These inductances and capacitances can be reduced, and the upper frequency limit extended, by reducing the physical dimensions of the tube structure and by making arrangements such that the inductances and capacitances are minimized. One disadvantage of a smaller structure is that it will have less heat-dissipating ability. In general, tubes for very high frequencies are quite small in comparison with tubes for lower frequencies.

Heating.—Temperature rise is a factor limiting the capacity of a tube. As the frequency of operation is increased, (*a*) the skin-effect losses increase; (*b*) larger currents flow because the reactances of the various capacitances are less, and these larger currents cause increased I^2R losses; (*c*) the electromagnetic radiation of power by the tube and circuit is greater; and (*d*) the dielectric hysteresis losses (page 53) in the tube, insulating material, and tube base are

greater at higher frequencies. Heating effects are reduced by using large conductors of good conductivity, by using special arrangements for bringing out the leads so that strong high-frequency electric fields are avoided, by using low-loss ceramic tube bases, or by eliminating the bases entirely.

Oscillators Using Resonant Transmission Lines.—For very high frequencies the inductance and capacitance required become very small. Also, because the losses in conventional coils and condensers increase with frequency, it is difficult to obtain tuned circuits with high Q. Quartz crystal control is difficult or impossible to apply at very high frequencies. For these and other reasons it becomes more satisfactory to use resonant transmission lines to control the frequency of operation. Such lines sometimes are called **Lecher wires.**

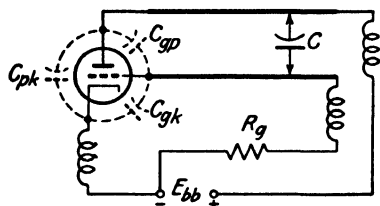



FIG. 210.—A simple oscillator using a resonant line to control the frequency. The coils shown are radio-frequency "chokes." These keep high-frequency currents from flowing in the leads and also reduce undesired radiation.

Transmission lines were discussed in Chap. V, where it was shown that at certain frequencies the lines behaved as tuned parallel inductance-capacitance circuits. If the line losses are small (and this can be achieved), these lines act as sharply tuned circuits, and are therefore well adapted to hold the frequency of an oscillator at a fixed value.

As explained in Chap. V, the input impedance of an open-circuited line (or coaxial cable) less than a quarter wavelength long is pure capacitive reactance; also the input impedance of a short-circuited line less than a quarter wavelength long is pure inductive reactance. Thus two such sections joined together and driven (by applying the alternating signal voltage) between the two ends will behave like a tuned parallel circuit composed of a condenser and a coil. Oscillators employing resonant lines (or resonant coaxial-cable sections) are used extensively at very high frequencies, and many different types have been developed. Two typical circuits will be described.

A simple oscillator using a resonant line to control the frequency is shown in Fig. 210. The condenser C can be slid along the two parallel conductors forming the transmission line. The coils are

radio-frequency chokes to isolate the radio-frequency energy from the direct-current power source. Often a second condenser is placed a quarter wavelength to the right of condenser C. The condensers essentially are short circuits at the very high radio frequencies generated. When the tube oscillates, the grid is driven positive, and the direct-current component flowing through R_g produces a direct-voltage drop that biases the tube. The con-

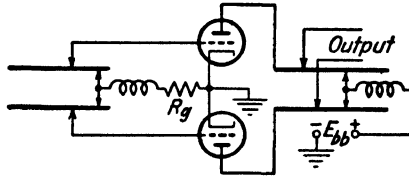
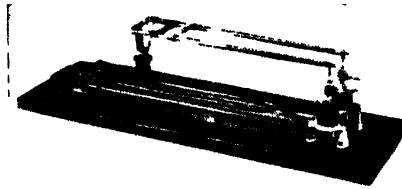



FIG. 211.—An oscillator using resonant lines in both the plate and grid circuits.

A small resonant-line oscillator using the two triodes at the right. This oscillator will produce about 25 watts of power at 300 megacycles.

denser C, which places a radio-frequency short circuit on the line, is adjusted to a position such that the length of line to the left is inductive. This section is then in parallel with capacitor C_{gp}, the grid-plate interelectrode capacitance. This provides the equivalent of a parallel tuned plate circuit that fixes the frequency of oscillation. The interelectrode capacitance C_{pk} from plate to cathode is in series with C_{gk} from grid to cathode, and these form a capacitive voltage divider, much as in the Colpitts oscillator. By means of this, signal voltage and power are fed into the grid circuit from the plate circuit and oscillations are produced.

An oscillator using resonant lines in both the plate and grid circuits is shown in Fig. 211. The plate and grid lines are tuned to the equivalent of parallel resonance by moving the shorting bars. In each instance resonance is obtained with the interelectrode tube capacitance, as explained in the preceding paragraph. The feed-

back of signal voltage and power from the plate circuit to the grid circuit is through the grid-plate capacitance. The position of the output taps depends on the impedance of the connected load. This is a push-pull oscillator, is balanced to ground, and is well suited for working into an open-wire transmission line. With ordinary small power-output tubes, frequencies up to about 60,000,000 cycles and power outputs of many watts can be obtained. The circuit can be modified to use pentodes or beam-power tubes. When special high-frequency tubes are used, frequencies of several hundred megacycles readily are obtained.

Positive-grid Triode Oscillators.—The oscillators thus far considered operated in the usual manner with the grids biased negatively. Oscillators for generating ultrahigh frequencies (300 to 3000 megacycles, or 100 to 10 centimeters) have been developed in which the *plate* is negative and the *grid* is positive. Such oscillators, developing wavelengths of but a few centimeters, were used in communication systems as early as 1930. An explanation of these positive-grid oscillators is as follows.

When certain triodes are operated with the grid *positive* and the plate *negative*, some of the electrons attracted to the grid pass between the grid wires. Their kinetic energy carries them on toward the plate. They never reach it, because both they, and the plate, are negative. They are attracted back toward the positive grid. Some of the electrons will strike the grid wires, but others will pass through and on toward the cathode, which is negative with respect to the positive grid. The electrons soon will be brought to rest and will return toward the positive grid, where some will be removed and others will pass on toward the negative plate.

These electron oscillations by some of the electrons cause an ultrahigh-frequency voltage between the tube electrodes. For instance, when negative electrons approach the negative plate, the approaching electrons repel electrons from the plate, making the plate less negative (or more positive) than before. The frequency is determined by the tube construction and the electrode voltages. The efficiency and output are low because most of the electrons strike the positive grid and do not oscillate. These are called **Barkhausen oscillators.** Similar devices in which Lecher wires are used to regulate the frequency have been developed. These are known as **Gill-Morrell oscillators.** These oscillators have been replaced now to a great extent by cavity oscillators,

using somewhat conventional tubes, and by cavity oscillators of the magnetron or klystron types.

The Cavity Resonator.—The devices for producing ultrahigh frequencies and superhigh frequencies (page 372) employ resonant cavities instead of resonant transmission lines for fixing the frequency of oscillation. The elementary theory of these cavities can be developed from ordinary transmission-line theory in the following manner.

If a short-circuited transmission line is slightly less than a quarter wavelength long, the input impedance will be pure inductive reactance (page 154). If the frequency is increased so that the line is exactly one-quarter wavelength long, the input impedance will be a high value of pure resistance. If the frequency further is increased, the input impedance will be pure capacitive reactance. These are the same characteristics as exhibited by a parallel tuned resonant circuit composed of a coil and a condenser. Since such tuned circuits can be used to establish the frequency of oscillation, it follows that quarter-wave resonant lines can be used to determine the frequency of oscillation.

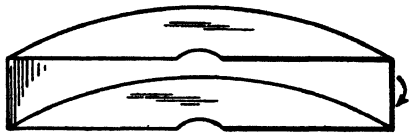



FIG. 212.—Moving a short-circuited transmission line as indicated by the arrow "generates" a cavity.

But radiation from a resonant line is great at ultrahigh and superhigh frequencies. Also, a resonant line for such frequencies becomes a small loop a few centimeters long. Hence, it becomes more practicable to use a resonant cavity instead of a line. In a sense, a resonant cavity is but many resonant lines with their "sides" touching. This is illustrated by Fig. 212. Moving the short-circuited transmission line, as indicated, "generates," geometrically speaking, a cavity. It was mentioned earlier in this chapter that a resonant transmission line was better than a tuned circuit consisting of a coil and a condenser, because the losses in the line were less, and a circuit with a higher Q was possible. Now at very high frequencies, and particularly at ultrahigh and superhigh frequencies, the currents flow only in an extremely thin surface layer, so thin in fact, that silver plating the surface is very effective in reducing the losses. Since a cavity has a large inner surface, it has lower losses and a higher Q than a line. For these and other reasons, such as the fact that little radiation occurs,

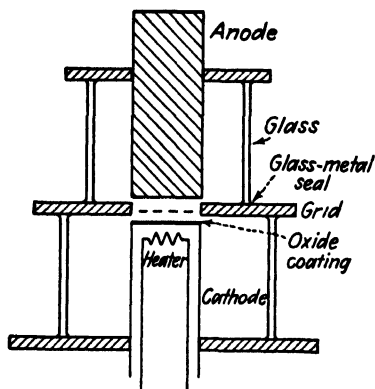



FIG. 213.—Cross section of a disk-seal or "lighthouse" tube. Because of the peculiar construction the electrode spacing can be kept to a minimum, thus reducing transit time.

cavities are largely used with modern oscillators in the ultrahigh and superhigh frequency regions.

The Disk, or "Lighthouse," Tube Oscillator.—Certain of the difficulties encountered in the operation at very high radio frequencies of ordinary vacuum tubes were discussed elsewhere in this chapter. Among these were the transit time and high interelectrode capacitance. About 1945, a tube was introduced with a different principle of construction. The electrodes of this tube are in the form of disks insulated from each other by glass cylinders or rings. Although triodes, tetrodes, and pentodes are available, this discussion will be confined to the triode because of its relative simplicity.

The basic principle of the disk tube, or disk-seal tube, or "lighthouse" tube (as it is popularly called because of its shape), is shown in Fig. 213. The portions of the electrodes useful electronically are parallel disks. With this construction close spacing and short transit time are possible, massive heat-dissipating electrodes with less inductance and capacitance are used, and the tube is particularly well adapted for use with cavities.

A disk or "lighthouse" tube. *General Electric Co.)*

The elementary theory of the cavity oscillator and the disk tube can be explained in the usual manner. Thus the transmission-line oscillator of Fig. 210 was shown to be a form of the Colpitts oscillator of Fig. 204, having tuning only in the plate circuit. The oscillator of Fig. 211 used tuning in both the grid-cathode and plate-

cathode circuits. In these two oscillators the tubes were operated in the usual manner with the cathode circuit grounded.

In Fig. 214, the grid circuit is grounded, and tuning is placed between the cathode and grid and the plate and grid, with signal feedback occurring through the plate to cathode interelectrode capacitance. This is just a modification of conventional oscillator circuits. The capacitances and inductances shown in Fig. 214 constitute parallel tuned resonant circuits, and these effects may be provided by actual inductors and capacitors, by resonant transmission lines, or by resonant cavities.

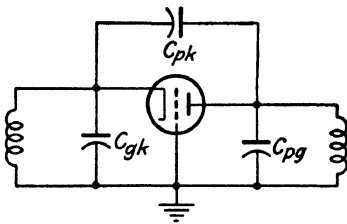



FIG. 214.—Simplified circuit of an oscillator with grounded grid.

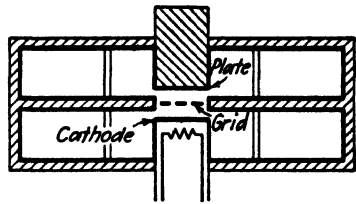



FIG. 215.—Showing the adaptation of the disk-seal or "lighthouse" tube to use as a cavity resonator. The cavities are outlined by the heavy lines.

The use of a disk, or lighthouse, tube in a cavity oscillator is shown in Fig. 215. The boundary of the cavity has been heavily outlined for clarity. Of course the mouth of this cavity is small compared with that of Fig. 212, but the basic action is the same. As in Fig. 214, the grid is common, or better yet it serves to separate one oscillating cavity from the other.[1] In considering this viewpoint it should be remembered that at the extremely high frequencies involved the currents flow only on the thin inner surface layers of the cavities.

Of course the circuit of Fig. 215 is schematic only, and it does not include many details, such as tuning arrangements, direct-current supply, and method of drawing ultrahigh-frequency power from the cavity oscillator. For these details the article by McArthur[1] will be found useful. Power is drawn from the cavity by coupling in a coaxial cable with a loop of wire, as will be explained for the two devices considered in the following sections.

[1] McArthur, E. D., Disk-seal Tubes, *Electronics*, Vol. 18, No. 2, February, 1945.

In considering the arrangement of Fig. 215, it is important to note that the tube and cavity merge into a common structure. This is typical of the devices used at ultrahigh and superhigh frequencies. It is difficult or impossible to define where the tube stops and the circuit begins.

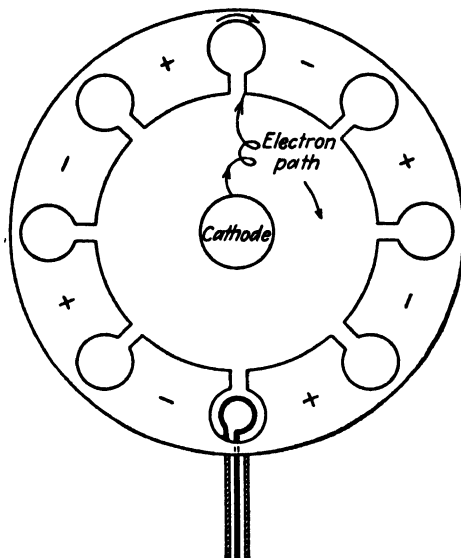



FIG. 216.—In the magnetron, the electrons from the cathode are attracted to the positive anode by a voltage (not indicated) impressed between the cathode and anode. Because of the magnetic field (not shown) passing between the cathode and anode, the electrons are caused to follow various and complicated paths. *One possible path* is as shown, and many others exist. In fact, the electron paths may be likened to the spokes of a wheel extending out from the cathode, and these spokes may be thought to rotate as the arrow indicates. Such action would produce the instantaneous potentials indicated and would produce *conventional* current flow as indicated for *one* (the upper) cavity. (*Adapted from Western Electric Oscillator, July,* 1946.)

The Magnetron.—This device operates because of the *combined* action of magnetic and electric fields. The magnetron consists essentially of a thermionic cathode at the center of a cylindrical positive anode. The magnetic field is produced by an electromagnet or a permanent magnet *external* to the tube. The magnetic lines of force pass through the region between the cathode and anode. These lines of force are *parallel* to the axis of the cathode and the anode that surrounds it. Electrons emitted by the hot cathode are attracted by the positive anode and flow toward it.

Because these electrons travel at right angles to the magnetic field, their paths are not straight, but are curved and become quite difficult to determine. In fact, the operation of a magnetron still is under study, and the following should be regarded as an approximate explanation.

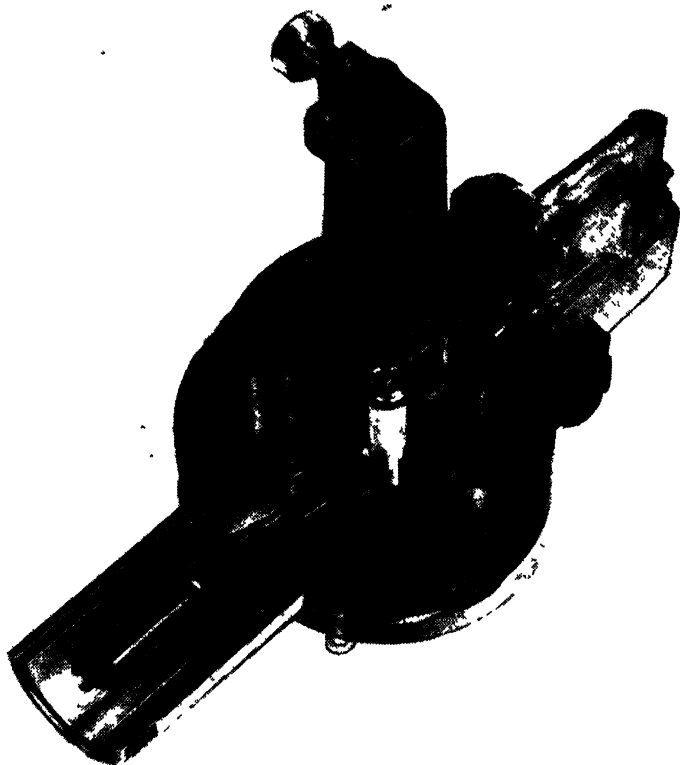
The magnetron has been in use for many years, and in various forms. The schematic diagram of a magnetron is shown in Fig. 216. For simplicity the magnetic field is omitted. The illustration shown is an "end view," looking down along the cathode and the walls of the anode. The poles of the magnet are assumed to be above and below the page, and the magnetic lines of force are between the cathode and anode, and parallel to the cathode and anode surfaces. The cathode is a large indirectly heated oxide-coated structure.

The anode is (relatively speaking) a massive copper electrode containing cavities. These constitute tuned oscillating circuits. The paths of the emitted electrons as they travel to the positive anode are many and varied. Some electrons are forced back to the cathode by the magnetic field; others more or less spiral toward the anode. In considering why the moving electrons are influenced by a magnetic field, it should be remembered that moving electrons constitute a current flow and that a current (as in a wire) is deflected by a magnetic field. *One possible path* is as indicated, but it should be remembered that a multitude of other paths exist and are being followed by electrons. Also, the paths *are revolving* rapidly somewhat as the spokes of a wheel as indicated by the arrow.[1]

As a result of the electrons striking the various portions of the anode, one side of each cavity is made more negative than the other side at a given instant. This causes a difference of potential between the openings of the cavities, and it causes current flow along the walls of the cavities. Because of the rapid motion of the electron paths around the anode, as indicated by the arrow, the polarities of the regions between the cavities and the polarities at the cavity openings alternate very rapidly, and frequencies of billions of cycles are produced. A coaxial cable coupled with a loop to one of the cavities will draw power from the magnetron. The currents in the cavity walls induce corresponding alternating voltages in

[1] This discussion and Fig. 216 are based on an article by J. C. Johnson, Cavity Magnetrons. This article appeared in the July, 1946, issue of the *Oscillator*, a publication of the Western Electric Company. Another type of magnetron is made of segments and has cavities that are somewhat triangular.

the loop. Usually a coupling loop is placed in one cavity only. Sometimes the output is directly into a hollow metallic tube called a **wave guide** instead of into a coaxial cable.

In a typical magnetron, as used in a radar system, the direct voltage of perhaps 10,000 volts is applied periodically between the

A cavity magnetron that is tunable and that can produce a variable frequency. The range covered is from 900 to 970 megacycles. (*Western Electric Co.*)

cathode and anode. Such a magnetron accordingly produces periodic pulses of extremely high-frequency alternating-current power. Because of the short duration of these pulses the instantaneous voltage, current, and *peak* power output may be very high, of the order of 1000 kilowatts. Frequencies of about 10,000 megacycles and corresponding wavelengths of 3 centimeters are produced readily.

The Klystron.—Although this device will operate as an amplifier, and in other ways, the only application that will be considered is its use as an oscillator, or generator, of ultrahigh frequencies. Furthermore, it exists in many forms, but only one of the most easily understood varieties will be discussed.

The essential parts of a klystron are shown in Fig. 217. Electrons are emitted by the thermionic oxide-coated cathode. The number of electrons leaving the cathode region and flowing down the tube is controlled by a control grid which is positive with respect to the cathode and which acts as an accelerator.

The electrons are formed into a beam, or ray, and travel toward the two gridlike openings of a resonating cavity known as the **buncher.** Some of the electrons strike the grids of the buncher, but others pass through the **drift space** and pass toward the gridlike structure of a second resonating cavity called the **catcher.** Here again some of the electrons strike the catcher grid, but the remainder pass on through to the collector, and then back to the cathode.

In the klystron oscillator, or generator, the two cavities are connected by a coaxial cable. The ends of the cable are terminated in small loops, and these take energy from, or deliver energy to, the cavities. In this connection it should be remembered that the distant end of a cavity is similar to the short circuit at the distant end of a transmission line, and that when a cavity is excited by high-frequency electric energy large currents are flowing in the distant end of the cavity where the pickup loops are located.

As shown in Fig. 217, the exposed metal cavities are at ground potential, and the enclosed electron-emitting cathode is at a high negative potential. Grounding the cavities is a safety precaution. Of course the control grid and the cavities still are positive with respect to the cathode.

To explain in an elementary way why the klystron oscillator generates alternating-current power of extremely high frequencies, assume that a group of electrons from the cathode approach the first grid of the buncher. As this occurs, the oncoming negative electrons will drive negative electrons out of the first grid of the buncher and around the walls of the buncher cavity into the second grid of the buncher. This causes current flow in the walls of the buncher and a difference of potential between the two grids of the buncher. When the moving electrons under consideration are mid-

way between the two grids of the buncher, equal forces are exerted on the two grids and no difference of potential exists between them. When the electrons reach the vicinity of the second grid of the buncher, the electrons repel electrons from this grid, and cause a difference of potential across the buncher cavity. If a concen-

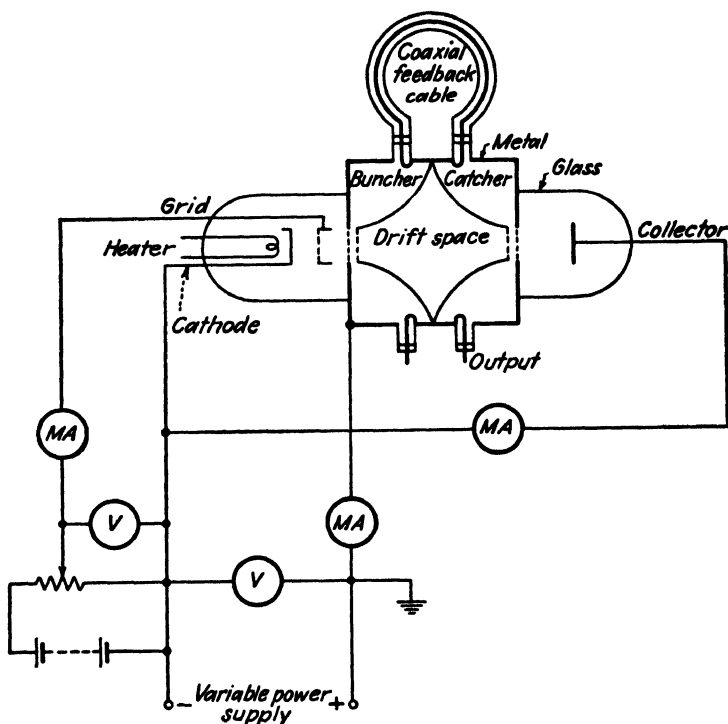



FIG. 217.—Showing the arrangement of the electrodes in one of the early klystrons, and the external test circuit for studying its operation.

trated group of electrons instead of a small group passed down the axis of the tube, the action would be intensified, a greater difference of potential would be created at the cavity opening, and larger currents would flow in the cavity walls. Similar results would exist when the concentrated group of electrons passed through the catcher.

An electron beam seldom is perfectly uniform. In the first place it is composed of particles (electrons) and these have random motions at least to some extent. Some electrons are traveling faster than others and tend to "catch up" with those ahead. This

bunching action can be increased if an alternating voltage is caused to exist across the grids at the opening of the buncher cavity. Such an alternating voltage is made to exist at the buncher cavity by the alternating-current energy fed back from the catcher to the buncher through the coaxial cable.

Thus the random nature of the electron beam first causes a feeble difference of potential between the buncher and catcher grids, and the signal fed back from the catcher to the buncher intensifies the action by further accelerating or retarding the oncoming electrons so that they "bunch up" in the drift space. Thus oscillations build up in the klystron much as in an ordinary vacuum-tube oscillator where signal is fed back from the plate to grid circuit.

A cut-away view of a klystron of a type somewhat like the one shown in Fig. 217. (*Sperry Gyroscope Co.*)

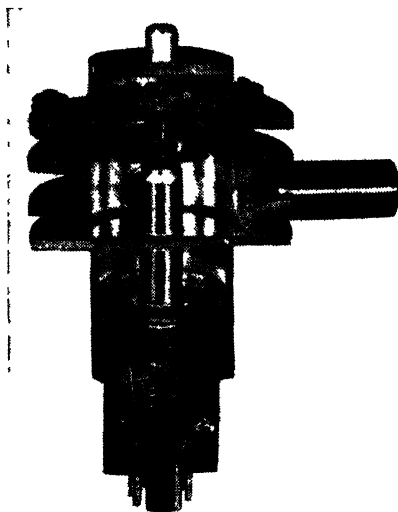
Of course, the distances between the grids, the length of the drift space, the length of the feed-back coaxial cable, and the potential (and resulting electron velocities) must be correct for oscillations to occur. Some adjustment of frequency is possible by changing the applied voltage and by altering slightly the spacings of the resonant cavities. A klystron oscillator, such as the one discussed here, works well at about 3,000 megacycles, producing 10-centimeter waves. Outputs of several hundred watts are possible with air cooling. It is again mentioned that this is but one of many klystrons. The frequency of the so-called **reflex klystron** can be varied over a wide range by changing the applied voltage.

SUMMARY

Vacuum-tube oscillators are used extensively as sources of radio-frequency alternating-current electric energy. Both ordinary and special tubes are used. The usual oscillator is merely an amplifier with feedback such that the oscillations are sustained. The vacuum-tube oscillator converts direct-current power to alternating-current power. Oscillators can be made to produce frequencies in the audio-frequency range (and below) and also up to frequencies of many billions of cycles.

Audio-frequency oscillators may be specified arbitrarily as those producing frequencies from 25 to 25,000 cycles, based on the fact that these frequencies are audible to the human ear. Audio-frequency oscillators include three important types: (*a*) the *L-C* oscillator; (*b*) the *R-C* oscillator, and (*c*) the beat-frequency oscillator.

The frequency of the *L-C* oscillator, which uses the ordinary parallel tuned circuit of inductance *L* and capacitance *C* to determine the frequency of oscillation, is given by the approximate equation

$$f = \frac{1}{2\pi \sqrt{LC}}.$$

A modern reflex klystron that operates in the 1825- to 2100-megacycle band and that has an electron tuning range of 15 megacycles. This tube has a signal power output of about 5 watts. One use is in the so-called microwave relay stations where the signal is amplified in a point-to-point radio system. (*Sperry Gyroscope Co.*)

The tubes of a typical audio-frequency oscillator operate in class A1 so that the distortion will be low. Usually a buffer stage is inserted between the oscillating tube and the output circuit.

The frequency of the *R-C* oscillator, which is essentially a resistance-coupled amplifier with controlled feedback, is given by the approximate equation

$$f = \frac{1}{2\pi \sqrt{R_1 R_2 C_1 C_2}}.$$

where R_1, R_2, C_1, and C_2 make up the feed-back circuit.

Radio-frequency oscillators may arbitrarily be considered to include all those operating above the audio-frequency band, and would accordingly be

from about 25,000 to 30,000,000,000 cycles. A further classification is used, as in the following table:

Classification	Abbreviations	Megacycles	Centimeters
High frequency............	H.F.	3—30	10,000—1000
Very high frequency.......	V.H.F.	30—300	1000—100
Ultrahigh frequency.......	U.H.F.	300—3000	100—10
Superhigh frequency.......	S.H.F.	3000—30,000	10—1

The frequency of a radio-frequency oscillator often is held constant by a parallel tuned circuit composed of inductance L and capacitance C, and the equation previously given applies. A method widely used in radio transmitters is to fix the frequency of operation with a quartz crystal. Quartz crystals can be made to oscillate with a basic frequency up to about 15 megacycles. For obtaining control of higher frequencies, frequency multipliers (page 349) can be used. A crystal has been announced that will oscillate in such a way that direct crystal control up to 50 megacycles is possible.

Radio-frequency oscillators usually operate in class C, with the tuned parallel circuit passing the fundamental and rejecting the harmonics. Intermittent operation may occur, but the consequences are not serious. Blocking, however, may result in serious damage to the tube and to equipment.

Two common radio-frequency oscillators use the Hartley and the Colpitts circuits. These differ in the manner in which the feedback is obtained. Generally speaking, an oscillator may have a tuned circuit in the plate lead, in the grid lead, or in both the plate and grid circuits.

Radio-frequency oscillators may use triodes, tetrodes, pentodes, or beam-power tubes. Quartz-crystal control can be used with each of these. Electron-coupled oscillators avoid the necessity of a separate buffer stage.

At about 15 megacycles, the output of an oscillator (or amplifier) using ordinary vacuum tubes begins to decrease. If minor changes are made, the usual type of construction can be used up to about 60 megacycles before the output drops off appreciably. Above this frequency, special tube structures are used. In general, these incorporate changes to reduce the transit time and the electrode, lead, and base capacitances and inductances.

At very high and ultrahigh frequencies resonant sections of transmission lines are used to control the frequency of oscillation. These sections behave like parallel tuned circuits.

Ultrahigh frequencies can be generated by positive grid oscillators, but their outputs are weak. More satisfactory results are obtained with cavity oscillators using the disk or "lighthouse" tube or with the magnetron or klystron. The disk tube is, electronically speaking, a conventional tube with a cathode, grid (or grids), and a plate. These are somewhat like disks, however. The magnetron and klystron are special devices, operating on principles different from the ordinary vacuum tube.

In the magnetron, a magnetic field is used to curve the paths of the electrons and to cause the paths to rotate so that the electrons are swept across cavities, causing them to oscillate. In the klystron oscillator a beam of electrons

causes a difference of potential at the openings of cavities. Through feedback, the bunching and the magnitudes of the induced voltages are increased. Cavities behave much like tuned transmission lines and tuned parallel circuits in controlling the frequency of oscillation.

REVIEW QUESTIONS

1. Why may an oscillator be considered to be an amplifier with feedback?
2. Discuss positive and negative resistance, and give examples.
3. Enumerate the ranges covered by audio-frequency and radio-frequency oscillators.
4. What is meant by an *L-C* oscillator, and by an *R-C* oscillator?
5. Why is a parallel tuned circuit used in oscillators?
6. Is there any relationship between the parallel tuned circuit as used in an oscillator and as used in an amplifier?
7. Why is it possible to use Eq. (117), when this is the equation for series resonance?
8. Why are the tubes of audio oscillators usually operated in class A, while in radio oscillators they usually are operated in class C?
9. What is the cause and effect of intermittent operation?
10. What is the cause and effect of blocking?
11. For what purposes are high-frequency oscillators used in industry (as distinguished from communication)?
12. What is the difference between the Hartley and the Colpitts oscillators?
13. Discuss the principle of operation of the quartz-crystal oscillator.
14. What is a buffer, and why is it used?
15. Is a buffer necessary in an electron-coupled oscillator?
16. Why are parallel tuned circuits instead of series tuned circuits used in the plate lead of radio amplifiers and oscillators?
17. Discuss some of the reasons why ordinary vacuum tubes are not entirely satisfactory at high radio frequencies.
18. Discuss the points of similarity between tuned parallel circuits and resonant transmission lines.
19. On page 389 it states that the oscillator of Fig. 210 is similar to a Colpitts oscillator. Explain.
20. Briefly discuss the operation of a positive-grid oscillator. Are they of present practical importance?
21. In what important electrical respects are cavities similar to resonant lines?
22. Name several advantages of cavities over resonant lines.
23. What are the important electrical advantages of the disk-tube construction?
24. Briefly explain from an elementary viewpoint how a magnetron operates as a high-frequency generator.
25. Briefly explain the principle of operation of the klystron oscillator.

PROBLEMS

1. The choke coil used in the plate circuit of the audio-frequency oscillator of Fig. 201 has an inductance of 15 henrys and a direct-current resistance of

350 ohms. Assume that these values remain fixed, and calculate the plate-load impedance offered by this coil at 100, 1000, and 10,000 cycles.

2. Both tubes used in the circuit of Fig. 201 have the coefficients $\mu = 13.8$ and $r_p = 12{,}000$ ohms. The bias is -5 volts, and $R_L = 30{,}000$o hms. Calculate the value of R, and calculate the approximate maximum alternating voltage output of the oscillator. Remember that the oscillator tube operates essentially in class A.

3. For the oscillator of Fig. 201, what values of L_2 and C should be used for generating frequencies of 100, 1000, and 10,000 cycles?

4. A Hartley oscillator is to be constructed as in Fig. 204*a* for a frequency of 1,000,000 cycles. A 500-micromicrofarad variable air condenser is available and is to be used as condenser C. It is assumed that L_2 will be four times L_1. Calculate the approximate values of L_2 and L_1, and express the answer in microhenrys. In your opinion will these values of C, L_2, and L_1 be satisfactory for a general-purpose oscillator?

5. Consult a handbook for the equations that apply, and design the coil L_1-L_2.

6. A tube is being operated at 100 megacycles. The time required for an electron to pass from the cathode to the plate is about 0.001 microsecond. What phase angle will the voltage between the cathode and plate have gone through during the time required for the electron to pass from cathode to plate?

7. A quarter-wave resonant line is to be used with a 120-megacycle oscillator. What must be the length of the line in inches?

8. For the particular cavity of Fig. 212, the cavity "generated" by revolving a quarter-wave line will not be in resonance at the same frequency as the line. For this particular shape, $r = 0.38\lambda$ for resonance, where r is the diameter of the cavity, expressed in the same units as the wave length λ (see reference given in footnote, page 393). What should be the radius in centimeters for a cavity to be resonant at 4000 megacycles? Do you believe that this radius would hold if the cavity openings were brought close together as in Fig. 215? Explain.

9. A high-frequency oscillator is to be used to heat and harden the surface of a metal shaft. Should the magnetic or electric field be used for this purpose? Why? Draw a sketch indicating how the high-frequency energy should be applied to the shaft.

10. A high-frequency oscillator is to be used for the quick thawing of frozen food that exists in small, rectangular, waxed cardboard containers. Should the magnetic or electric field be used for this purpose? Why? Draw a sketch indicating how the high-frequency energy should be applied to the frozen food.

CHAPTER XI

MODULATION AND DEMODULATION

The purpose of a radio system is to transmit speech and music, or the code signals used in telegraphy, or the electric impulses from a television camera. The frequencies at which these signals originate are comparatively low with respect to the radio frequencies at which they are transmitted through space.

These matters were discussed on page 42, and the over-all picture of a radio system was presented in Fig. 25. It is of importance to note that the original signals are moved up to radio frequencies by the radio transmitter and that, after passing through space, the signals are moved back down to their original frequencies.

The process by which the original signals are moved up to radio frequencies is known as **modulation.** The process by which the radio-frequency signals are moved back down to their original frequencies is known as **demodulation.** Other names, such as **detection,** often are used, particularly in radio.

From the standpoint of the *results* accomplished, modulation and demodulation are opposite processes because modulation moves the signal frequencies up to higher values, and demodulation, or detection, moves the frequencies back down to their original values. From the standpoint of the *processes* used, modulation and demodulation may be regarded as identical.

Although the basic processes by which modulation and demodulation are achieved are identical in principle, the signal strengths involved, and the circuit details, are different. For this reason, modulation and demodulation will be considered separately.

Types of Modulation.—As defined in the standards,[1] modulation "is the process whereby the amplitude (or other characteristic) of a wave is varied as a function of the instantaneous value of another wave." The first wave, which is usually a single-frequency wave, is called the **carrier wave;** the second wave is called the **modulating wave.** In radio systems, the carrier wave is a radio-fre-

[1] American Standards of Electrical Terms, 1941.

quency wave generated by an oscillator, and the modulating wave is of relatively low frequency and consists of the speech, music, or other signals to be transmitted.

As mentioned in this definition, the carrier wave which is modulated by the signal wave to be transmitted *usually* is a single-frequency wave, that is, a pure sine wave. Radio systems are possible, however, in which the high-frequency wave to be modulated is not a sine wave, but consists of pulses. Most of this chapter will be devoted to the modulation of a sinusoidal carrier wave, because this method is the one largely used at present.

Before continuing, the word "carrier" should be discussed. In the early days of radio the viewpoint was held that the carrier wave that was modulated "carried" the signal wave to the distant radio-receiving station—hence, the term "carrier." This viewpoint is not correct; many of the most important radio systems suppress the carrier wave entirely or in part, and do not transmit the carrier-frequency, or transmit it at greatly reduced strength.

In the system of modulation using a sinusoidal carrier, the carrier and the modulating signal, such as the voice, are impressed simultaneously on the **modulator**, which is the device to effect or cause the process of modulation (see footnote, page 266). Modulation can be accomplished by acting on any *one* of the characteristics of the sinusoidal carrier wave. The basic equation for the instantaneous value of any sinusoidal voltage (such as a carrier wave) is

$$e = E_{max} \sin \theta \qquad \text{or} \qquad e = E_{max} \sin (2\pi ft + \phi), \qquad (120)$$

where e is the voltage at any instant, E_{max} is the maximum value of the voltage, and θ is the phase angle of the wave at the instant under consideration. The angle θ is composed of the angles $2\pi ft$ and ϕ, where f is the frequency, t is the time, and ϕ is an angle that may be varied at will. For the impressed carrier voltage, Eq. (120) becomes

$$e_c = E_{max} \sin (2\pi f_c t + \phi). \qquad (121)$$

It is of importance to note that f_c is the frequency of the carrier wave, and it is *maintained constant* by crystal control, or otherwise. The only way that the frequency of the carrier can be changed *is to alter the frequency at the source.* Of course this is undesired because a transmitter must be kept on the carrier frequency that is assigned to it.

Amplitude Modulation.—If the modulating speech, or other signal, wave when simultaneously impressed with the carrier on the modulator causes the value E_{max} of Eq. (121) to rise and fall in exact accordance with the modulating signal, then **amplitude modulation** results. This is the method that has had almost universal use in past years. With *no* modulating speech signal impressed, and only the sinusoidal carrier acting on the modulator, the output of the modulator (which may go direct to the transmitting antenna) will be a pure sine wave; that is, without modulation the output will be the carrier. When the modulating speech (or other) signal also is impressed on the modulator, in amplitude modulation the output of the modulator will rise and fall exactly as the modulating speech signal rises and falls. In other words, in amplitude modulation the characteristics of the modulating wave will be impressed on the *signal output* of the modulator. A word of warning is necessary at this point: In amplitude modulation, it is not the carrier-frequency component that rises and falls in accordance with the modulating signal; *it is the output of the modulator.* This output is a simple wave of carrier frequency only when the modulating signal is not impressed. When the modulating signal is impressed, the output of the modulator is a new and complex signal, composed of several frequency components among which may be, or may not be, the carrier frequency component, depending on the system of amplitude modulation used.

Angular, or Angle, Modulation.—If the angle ϕ of Eq. (121) is varied by the modulating signal, then **angular,** or **angle, modulation**[1] results. Angular, or angle, modulation is of two types, **frequency modulation** and **phase modulation.** These two methods are very similar in many details. They will be considered on page 439.

The method to be followed in this chapter will be to consider amplitude modulation and demodulation first, then frequency and phase modulation and demodulation, and the pulse system last.

Amplitude Modulation.—The amplitude modulation of a sinusoidal carrier wave by the usual means results in an output from the modulator that varies in amplitude both above and below the axis in accordance with the modulating signal (Fig. 218). For the purpose of analysis it is convenient to consider that the modulating

[1] Everitt, W. L., Frequency Modulation, *Electrical Engineering,* Vol. 59, No. 11, November, 1940.

signal is a pure sine wave, just as if a person whistled (with say, a pitch of 1000 cycles) into the microphone connected to a radio transmitter. If a 1000-cycle wave is used to modulate a carrier wave of higher frequency, the output wave of the modulator will appear as in Fig. 218. The envelope of this wave, as indicated by the dotted line, will be the same shape as the modulating signal. In Fig. 218, this envelope appears as a sine wave, because a 1000-cycle sinusoidal wave was used to modulate the carrier. If speech had been used, then the envelope would be that of the complex speech wave (page 15).

FIG. 218.—When a high-frequency carrier wave is amplitude modulated, the envelope of each side of the modulating wave is a replica of the modulating signal. In this figure the modulating signal is a sine wave, and hence the envelopes are sinusoidal. (The original modulating signal exists in the output of many modulators, but this is removed by the tuned circuits in radio and hence is not shown in this figure.)

Much confusion exists regarding the nature of a modulated wave. The modulated wave appears as the *solid* lines of Fig. 218. The dotted lines were *drawn in to produce the envelope*. In no sense of the word does the modulated wave of Fig. 218 indicate that the "carrier" is "carrying" the signal. In fact, Fig. 218 is not the carrier wave at all, it is a complex wave, created by the process of modulation, and it contains several frequency components (including the carrier).

In support of these statements, Fig. 219 has been included. This shows a portion of the wave of Fig. 218. Now if waves *b*, *c*, and *d* are carefully added point by point, it will be seen that the familiar modulated wave of Figs. 218 and 219*a* results. Thus it can be said that a modulated wave, such as Figs. 218 and 219*a*, is composed of several frequency components. These are *first*, a carrier component *b*, *second*, a lower side frequency *c*, and, *third*, an upper side frequency *d*.

A study of the shape of the envelope of a modulated wave is useful in many ways. For example, it is useful in a graphical study of vacuum-tube operation. Also, it shows the percentage modulation. Furthermore, if the shape of the envelope is not the same as the shape of the modulating signal wave, then undesired distortion has resulted. This type of distortion should not be confused with the desired distortion causing the creation of the side frequencies. Undesired distortion would result in distorted sounds when the radio signal was received and demodulated.

The Process of Amplitude Modulation.—It has been explained that in the process of amplitude modulation the radio-frequency carrier wave and the low-frequency signal wave are impressed simultaneously on a modulator and that a complex wave such as Fig. 218 results. From Fig. 219, it is seen that this wave consists of a carrier and two side frequencies.[1]

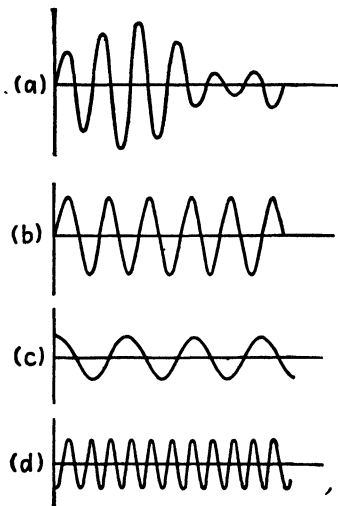



FIG. 219.—When a high-frequency carrier wave is amplitude modulated by a sine wave of lower frequency, a signal such as Fig. 218 and (*a*) above results. This modulated signal is composed of the carrier (*b*); a lower side frequency (*c*); and an upper side frequency (*d*). If the modulating signal is not sinusoidal, but is a complex wave such as speech, then the side frequencies become sidebands. (See statement in parentheses, Fig. 218.)

The question immediately arising is this: from where do the two side frequencies come? The radio-frequency carrier and the low-frequency modulating signal are impressed on the modulator, and it is not surprising to find these components in the output (unless a method is used in which they are suppressed). Before accounting for the presence of the side frequencies, it is of value to review the theory of rectification on page 221. As shown there, when a *single* frequency is impressed on a rectifier, the output wave *is distorted;* that is, the output wave does not look like the input wave, and distortion is defined (footnote, page 266) as a change in wave form. The rectified (distorted) wave was found to contain harmonics. Where did these harmonics originate? The only answer is that they were created in the process of nonlinear distortion (page 266), and from Fig. 128, page 222, these harmonics must exist to give the rectified wave its peculiar shape. Now modulation is a process of distortion, because if two pure sine waves are impressed *simultaneously* on a modulator, the peculiar pattern of Fig. 218 results. The conclusion must be drawn that the side frequencies *are created in the process* of nonlinear distortion known as "modulation."

[1] This is not a complete picture of modulation, because in many systems the modulating signal appears also in the output of the modulator. In a radio system, however, the output circuit is so tuned that the low-frequency modulating signal component does not exist in the output of the modulating stage.

The Nature of an Amplitude-modulated Wave.—A fundamental theorem applying to amplitude modulation is that when two sine waves of different frequencies are impressed simultaneously on a device that causes nonlinear distortion, sum-and-difference frequencies called "side frequencies" are created. Thus, if a 1,000,000-cycle carrier is modulated with a 1000-cycle wave (as by whistling into the microphone driving a radio transmitter), the output of the modulator will look like Fig. 219, and will contain the following components: *first*, a carrier, represented by *b*, having a frequency of 1,000,000 cycles; *second*, the lower side frequency, represented by *c*, having a frequency of 999,000 cycles; *third*, the upper side frequency, represented by *d*, having a frequency of 1,001,000 cycles; and *fourth*, the modulating signal of 1000 cycles that is assumed to have been suppressed and is not included in Figs. 218 and 219. Other unwanted distortion components also are produced, but these are suppressed and have been neglected.

Of course, many radio systems are built to transmit speech and music, and when a system is modulated with speech signals, the simple pattern of Fig. 218 does not result. Instead, the envelope of the output of the modulator is that of the speech signal variations, and **sidebands** instead of side frequencies are created in the modulating process. Thus suppose that a musical program covering the range of from 50 to 10,000 cycles is picked up by the microphone and is used to modulate the 1,000,000-cycle carrier. If, at a given instant, the program consisted of 50-cycle tones, the output of the modulator would contain components at 1,000,000 cycles, 999,950 cycles, and 1,000,050 cycles. If, at a given instant, the program consisted of 10,000-cycle tones, the output of the modulator would contain components at 1,000,000 cycles, 990,000 cycles, and 1,010,000 cycles. Thus the complex wave from a musical program composed of many tones and covering a range of 50 to 10,000 cycles if used to modulate a 1,000,000-cycle carrier would result in a modulated wave which contained sidebands, and which would contain (*a*) a carrier frequency component of 1,000,000 cycles, (*b*) a lower sideband of from 999,950 to 990,000 cycles, and (*c*) an upper sideband of from 1,000,050 to 1,010,000 cycles. This assumes, of course, that the system is designed so that the carrier is passed and the modulating signal is suppressed. This is the method used in radio-broadcast transmitters of the usual type.

If the magnitude of the impressed modulating signal equals that

of the modulated carrier, then the maximum values of the *modulated* output will be twice the magnitude of the *unmodulated* output, and the minimum values will be just zero. This is known as 100 per cent modulation, and the relations are shown in Fig. 220. The **percentage modulation** is the ratio of half the difference between the maximum and minimum amplitudes of a modulated wave to the average amplitude, expressed in per cent (see footnote, page 266). The percentage modulation is designated by m.

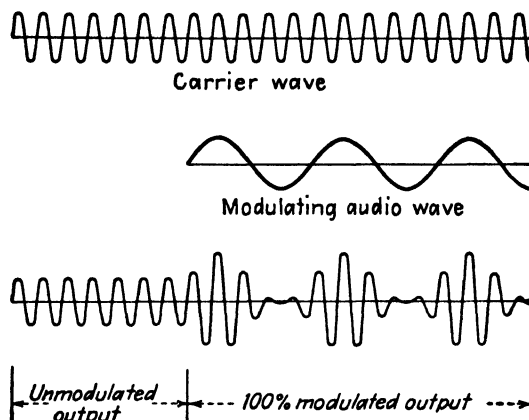



FIG. 220.—If the magnitudes of the modulated radio-frequency carrier wave and the modulating audio wave are the same, then the modulated wave will rise to twice the unmodulated output and will fall to zero. This is 100 per cent modulation.

As has been explained, amplitude-modulated waves are composed of a carrier and two sidebands. It is of interest to investigate the magnitudes of these. It follows from Fig. 219, that for 100 per cent modulation, the magnitude of *each* sideband component of an amplitude modulated wave is one-half that of the carrier component. These relations are shown in Fig. 221.

It is of interest to investigate the power (or energy) in each component of the modulated wave. Since, in general, the power $P = I^2R$ or $P = E^2/R$, it follows that the power in each component is as shown in Fig. 222; that is, the power in each sideband is one-fourth that of the carrier component. Or, expressed in another way, if the total power in the modulated wave is called 100 per cent, then $\frac{4}{6}$, or 66 $\frac{2}{3}$ per cent, of this total power is in the carrier component, and $\frac{1}{6}$ or 16$\frac{2}{3}$ per cent is in *each* sideband.

This is a very important matter. For instance, when modulated with a signal that is a pure sine wave, the *signal* output of a

standard amplitude-modulated radio-broadcast transmitter is *entirely* in the sidebands. The carrier component essentially is the same with or without modulation. Thus, 66⅔ per cent of the power is not conveying program, and 66⅔ per cent of the capacity of the expensive power-output stages is in a sense not being used.

Fig. 221.—Relative strengths of the carrier component I_c and the two sidebands I_s for 100 per cent modulation. This indicates the relative intensities or magnitudes of the components.

Ordinarily it is desired that the percentage modulation should approach 100 per cent on the peaks of the audio-frequency signals. The magnitude of the carrier remains essentially unchanged whether or not the circuit is modulated. If modulation approaches 100 per cent, then the sidebands, which contain the signal fluctuations and the information to be transmitted, will be strong. There is seldom (if ever) an advantage in using a large expensive tube and modulating it a small per cent; instead there usually is a marked saving in using a smaller tube and modulating it to approximately 100 per cent.

Fig. 222.—Relative power contents of the carrier P_c and sideband components P_s in a 100 per cent modulated wave.

As mentioned elsewhere, radio systems are in use that suppress the carrier largely or entirely and transmit the sidebands only. Such systems require a slightly more elaborate receiving set than that required for amplitude-modulated broadcast reception where the carrier is transmitted.

Methods of Amplitude Modulation.—This system has been used since the beginning of radio, and many methods of producing amplitude modulation have been developed. Also, amplitude modulation has been used in carrier-telephone systems that make possible the simultaneous transmission of many telephone messages over one pair of wires, and various methods of amplitude modulation have been developed for this purpose. No attempt will be made to cover all these methods; rather, the discussions that follow will include those systems of modulation of practical importance.

The literature on the subject of amplitude modulation is con-

fusing because little standardization of terms exists. Thus the audio-frequency amplifier that is used to drive (simultaneously with the impressed carrier wave) the portion of the radio transmitter that creates the sidebands often is called the modulator. If the standards (see footnote, page 266) are to be followed, the **modulator** is the device used to effect, or accomplish, the process of modulation. Hence, it would seem more logical to call the audio-frequency amplifier the **modulating amplifier,** and sometimes this is done in practice, and will be followed here. Also, the device used in radio to effect, or accomplish, the modulation often is called a modulated class C amplifier. The tube accomplishing the process of modulation often is operated as a class C amplifier, but this tube, and its associated circuit, actually is the device that produces the modulation, and is logically the *modulator.* There is a growing tendency to call it a **modulated amplifier.**

Amplitude Modulation by Grid and Plate Injection.—Three standard systems for amplitude modulation are shown in Fig. 223. In each drawing the tube and circuit at the left are the audio-frequency power amplifier, or modulating amplifier, and the tube and circuit at the right are the modulator which distorts the two simultaneously impressed audio and carrier waves, and which creates the desired sidebands, thereby raising or translating the low-frequency audio signals to radio frequencies for transmission.

The audio (or other) signals first are amplified in voltage amplifiers and then are impressed on the audio power-amplifying tube or tubes. These audio power-output tubes may be triodes, pentodes, or beam-power tubes, and may be operated in class A, AB, or B, just as for any audio power amplifier. They may be single tubes, or may be in push-pull. A push-pull class B amplifier is used often.

The carrier signal is produced by a crystal-controlled oscillator in most instances, although of course other oscillators may be used if permissible. In most instances operation is so rigidly specified that crystal control is required. The output of the crystal oscillator is amplified by class C power amplifiers before it is fed into the grid circuit of the modulator stage.

The radio-frequency choke is to prevent the radio frequencies from flowing into the modulating audio amplifier. Condensers C' are to prevent the L-C circuit from short-circuiting the plate

power supplies. Although grid-bias batteries E_{cc} and plate-supply batteries E_{bb} are indicated, these voltages usually would be obtained from rectifiers.

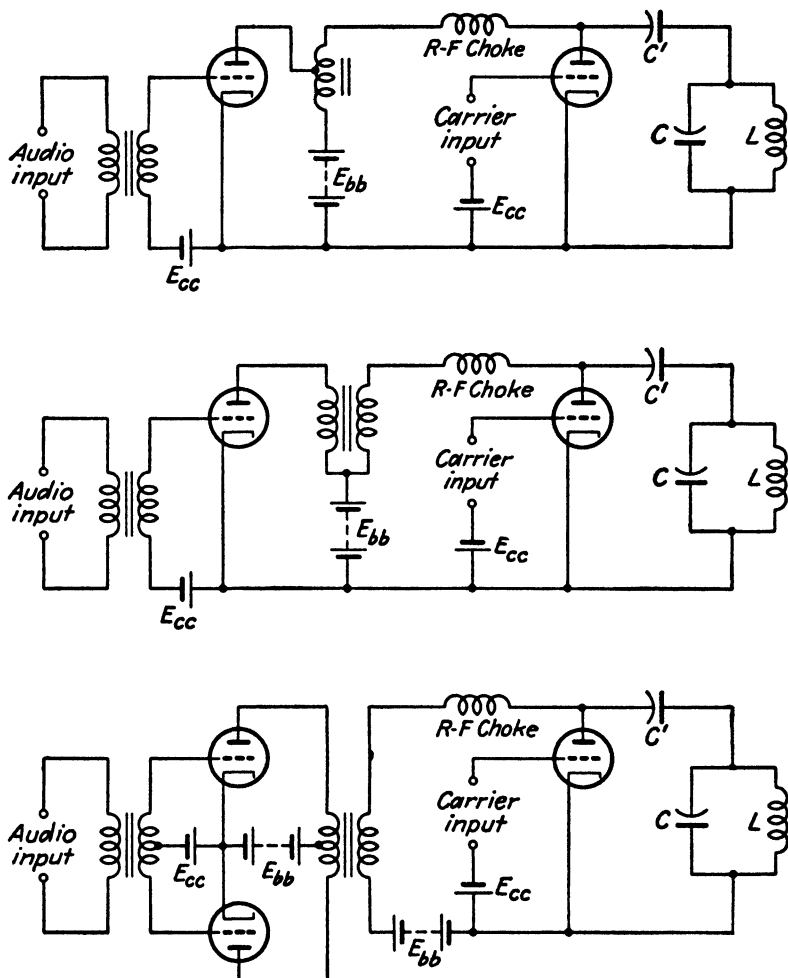



FIG. 223.—Circuits for the amplitude modulation of a class C amplifier by plate and grid injection. The output voltage appears across the tuned *L-C* circuit. An antenna, or the transmission line or coaxial cable feeding the antenna, usually is inductively coupled to the output coil *L*.

The main difference in the circuits of Fig. 223 is in the way the audio power amplifiers are coupled to the modulator circuit. In

the first circuit an audio-frequency choke is used as the load in the plate circuit of the audio power amplifier. Actually, the audio choke and the entire circuit to the right constitute the load on the audio power tube. The size of the choke that should be used will depend to some extent on the tubes employed, and is listed in some transformer catalogues. Often the choke is tapped, as indicated, so that the audio-signal voltage impressed on the second tube will be increased and 100 per cent modulation will be possible. It will be noted that the direct plate current for each tube flows through this coupling choke. In Fig. 223*b*, transformer coupling is used. This is far superior to choke coupling, because the audio amplifier can be properly matched to the circuit to the right, a better frequency response results, and 100 per cent modulation can be obtained. Transformer coupling also is used in Fig. 223*c*, and this basic circuit with class AB or class B push-pull audio amplification is very important.

Equivalent Circuit for Grid and Plate Injection Modulator.—As shown in Fig. 223, in each circuit the carrier wave is injected (or connected) into the grid circuit, and the audio-frequency signal is injected into the plate circuit. This system often is called **plate modulation.**

In each circuit of Fig. 223, the plate of the modulator tube has a direct voltage on it as furnished by E_{bb}. Also, the alternating output of the audio amplifier is in series with this direct voltage E_{bb}. Thus the circuits of Fig. 223 can be represented by the equivalent circuit of Fig. 224, which is convenient for explaining the way in which the modulation occurs.

Because the modulator tube is operated in class C, it is biased beyond cutoff, and if the carrier wave *only* is impressed on the tube, the shape of the current that will flow in the plate circuit will be as indicated by Fig. 224*a*. This current will contain a direct component, a fundamental component, and various harmonics created by the process of distortion (rectification).

If *no* alternating voltage, either modulating signal or carrier, is impressed on the modulator tube, the voltage between plate and cathode will be as shown by E_b of Fig. 224*b* and will be essentially the voltage of the source E_{bb}. When the low-frequency signal voltage E_s is injected, or impressed, into the circuit in series with the source, the voltage between plate and cathode will vary as shown by e_b.

The instantaneous variations of the plate voltage cause the modulated class C amplifier tube to be a poor amplifier when the instantaneous plate voltage is low, and a good amplifier when it is high. As a result, the amplification offered to the carrier wave impressed between grid and cathode will vary and the output plate current will be as indicated by Fig. 224c. The class C circuit should be so designed (page 343) that a linear relation exists between the magnitude of the output plate current and the magnitude of plate voltage.

Fig. 224.—Equivalent circuit for a modulated class C amplifier, and curves showing the operation. (a) Plate current with carrier only impressed. (b) Direct voltage E_b on the plate, and total voltage e_b when modulating signal is impressed. (c) Plate current when carrier voltage and modulating voltage both are impressed. (d) Voltage drop across tuned parallel circuit and current in tuning coil and condenser; this signal contains the carrier and two sidebands. These waves, and those in similar figures, are not actually peaked as shown, but are rounded like sine waves. They are as shown here for convenience in drawing.

Two waves of different frequencies are impressed simultaneously on the circuit, and distortion and modulation occur because of the fact that the output current depends on the *magnitude* of each wave. In the output plate current of Fig. 224c, many components will exist. Among these will be the low frequency modulating signal (the speech or program frequency), the carrier frequency, and the upper and lower sidebands that are created in the process of distortion. For a 1,000,000-cycle carrier, these will have the frequencies discussed on page 409.

It will be noted that, relatively speaking, the side frequencies (if a pure 1000-cycle signal is used to modulate the carrier), or the sidebands (if speech or a musical pro-

gram is used), have substantially the same frequency as the carrier. Thus if the tuned parallel circuit is designed so that it is not too sharply resonant (page 83), a large impedance will be offered by it to the carrier and the side frequencies (or sidebands), and a large voltage drop will exist across the tuned circuit for these *desired* components. Also, in accordance with the theory of the parallel circuit (page 80) large currents will flow in the *L-C* circuit for the desired components. The tuned circuit will offer but little impedance to the direct current, the speech component, and the various harmonics, and, hence, these do not appear as large voltage components across the tuned parallel circuit.

The shape of the voltage wave across the tuned circuit and the current wave in the *L-C* circuit will be as shown in Fig. 224*d*. This is for modulation with a single frequency, such as 1000 cycles, and for 100 per cent modulation. It is evident from Fig. 224*b* that for this to occur the peak of the alternating modulating signal voltage must equal the plate voltage on the class C tube. This is a detail of audio-frequency power-amplifier design and impedance matching (page 109). The impedance into which the audio-frequency modulating amplifier works is approximately a resistance of magnitude

$$Z = R = \frac{E_b}{I_b}, \tag{122}$$

where R will be in ohms, when E_b is the direct plate voltage in volts on the modulated class C amplifier tube and I_b is the direct current in amperes, both measured when the tube is not modulated.[1]

The power requirement for the class C carrier-frequency amplifier between the crystal oscillator and the modulator input must be known. Of course this will depend on the conditions of operation. The peak value of the carrier impressed on the grid of the class C modulator tube is such that the grid goes slightly positive with no modulating audio signal applied. If the grid bias on the modulator tube is known, then the carrier voltage also is known. Often the grid of the class C modulator tube is biased in the vicinity of two to three times plate-current cutoff. The grid of the class C modulator will draw power from the class C amplifier preceding it. The amount of power will depend on many conditions, because

[1] "The Radio Amateur's Handbook" contains much practical information that is very helpful in the design and construction of radio apparatus. This handbook is published by the American Radio Relay League.

during modulation the plate voltage of the class C tube varies, and hence the grid-current cutoff varies, as indicated in Fig. 225. The driving power required by typical triodes is listed in some tube catalogues. As an approximation, it is of the order of 10 per cent of the modulated plate-power output. The class C amplifier preceding the class C modulator must supply this power, and should have at least about 50 per cent overload capacity.

The power requirements for the class A or B audio power amplifier used to supply the modulating signal to the class C modulator is of special importance.[1] The magnitude of the carrier-frequency component in the output of the modulator is *the same* before and after modulation; that is, it is the same whether or not an audio-modulating signal is impressed. For this reason, all the power in the carrier wave must come from the source of plate power, E_{bb} of Fig. 223. (This neglects the input power to the grid circuit from the class C amplifier and crystal oscillator, as of course it should, because this power is lost in the grid circuit.) If the class C tube being modulated has an efficiency of about 70 per cent, and if the unmodulated carrier output is to be 1000 watts, then the direct-current input power will be about 1000/0.7 = 1430 watts. Since this does not change with modulation, the power that goes into the sidebands must come from the audio-frequency power amplifier that is modulating the class C modulator. For 100 per cent modulation, the power in the two sidebands will be one-half that in the carrier (Fig. 222). Hence, if the carrier is to contain 1000 watts, the sidebands must contain 500 watts. If the conversion of audio-frequency power to sideband radio-frequency power by the class C modulator also is 70 per cent, then the audio-frequency power amplifier must deliver 500/0.7 = 715 watts to the plate circuit of the class C tube being modulated. As mentioned elsewhere, the alternating signal voltage must be such that the plate is driven to zero (Fig. 224*b*). Not only is this a large amount of audio power, but also, audio-frequency amplifiers with tubes in class A

FIG. 225.—Showing the way in which the grid voltage and grid-current cutoff (dotted line) vary with modulation in a modulated class C amplifier.

[1] Here again, it should be mentioned that this audio-modulating power amplifier often is called the "modulator," but that in this text the class C stage that produces the distortion and creates the sidebands is called the "modulator."

or B are not nearly so efficient as a class C tube. Class B audio amplifiers are accordingly often used, but even with these, the power-handling capacity of the final audio tubes exceeds that of the tube being modulated. Thus in a typical installation, the physical size of the final audio class B tubes usually is greater than that of the class C tube being modulated when plate injection or plate modulation, as it usually is called, is used.

Two class C tubes with grid and plate injection, as discussed in this section, often are operated in push-pull (page 361). There are many advantages to this. One important advantage is that two small tubes with lower electrode voltages can be used to supply the same modulated output. Also, neutralization of tubes in push-pull is simple (page 360). Another reason is that undesired distortion is less.

Amplitude Modulation by Grid Injection.—In the type of modulation just considered, the carrier wave is impressed, or injected, into the grid circuit, and the modulating audio signal is injected into the plate circuit. In small transmitters particularly, it is desirable to inject both the carrier and audio signals into the grid circuit. This often is called "grid-bias modulation." This is a poor term, because the grid bias ordinarily is considered to be the constant direct-current voltage applied between grid and cathode. It is sometimes shortened to "grid modulation." This also is a poor term, because in true grid modulation, grid current flows, and the distortion and creation of sidebands occurs in the grid circuit.

Modulation, and the creation of sidebands, does not occur unless distortion results. For side frequencies, or sidebands, to be created, the circuit must cause nonlinear distortion. The idea that "beating two waves together" causes a "beat note" is entirely without foundation. Simultaneously impressing two waves (beating them together) causes beat notes (side frequencies) only when the circuit on which they are impressed causes nonlinear distortion. This misconception probably started because two tuning forks of different frequencies produce audible beats. The fact is that the audible beats do not exist in the air, but the beats (side frequencies) are created by the ear and the hearing process, which are nonlinear. It is easy to demonstrate this with two loudspeakers set on different frequencies, a microphone, a distortionless amplifier, and a wave analyzer. This system will measure the magnitude and frequency of each sound present. If the two loudspeakers are set on slightly

different frequencies, and if the sounds picked up by the microphone are studied with the wave analyzer, it will be found that no beats (side frequencies) exist.

To return to the subject of amplitude modulation by grid injection, a circuit for accomplishing this is shown in Fig. 226 of which there are many modifications. As is indicated in Fig. 226, both the radio-frequency carrier wave and the low-frequency modulating audio wave are injected into the grid circuit. The triode being modulated is operated as a class C amplifier.

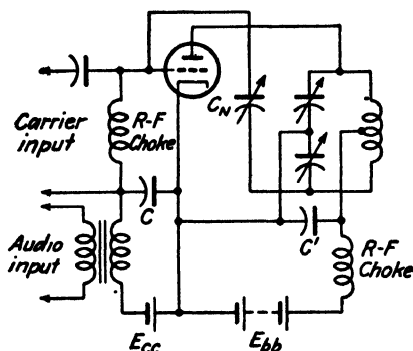



FIG. 226.—Simplified circuit for amplitude modulation by grid injection. Condenser C must have high reactance to the audio frequency and low reactance to the carrier. The modulated signal appears across the tuned parallel circuit in the plate lead.

With no modulating signal applied to the grid, but with the carrier voltage impressed, operation is from A to B of Fig. 227, and the plate current flows in "spurts," as for any class C amplifier (page 323). When the modulating speech signal voltage E_s is applied, the spurts of plate current rise and fall as in Fig. 227. For radio uses, the speech signals of a few thousand cycles are so low in frequency, as compared with the carrier of perhaps a million cycles or more, that the following viewpoint is useful.

It may be considered that the modulating speech, or progam, signal slowly (relatively speaking) varies the "bias" between the minimum negative and maximum negative values as indicated by E_s of Fig. 227. Thus, if the circuit, which is that of a class C amplifier, is analyzed at three points, operating data will be known. These three points are (a) at the actual grid-bias value with no modulation, (b) at the minimum "grid bias," when E_s is applied, and (c) at the maximum "grid bias."

The plate-current variations of Fig. 227 are the same as those of Fig. 224*c*, and the voltage across the tuned plate-load circuit *L-C*, and the currents flowing in the coil and condenser *L* and *C* will be as shown in Fig. 224*d*. This is the familiar amplitude-modulated wave containing the carrier and the sidebands that were created in the process of modulation.

Because the grid is driven positive, both the carrier-frequency amplifier and the audio-frequency modulating amplifier driving the

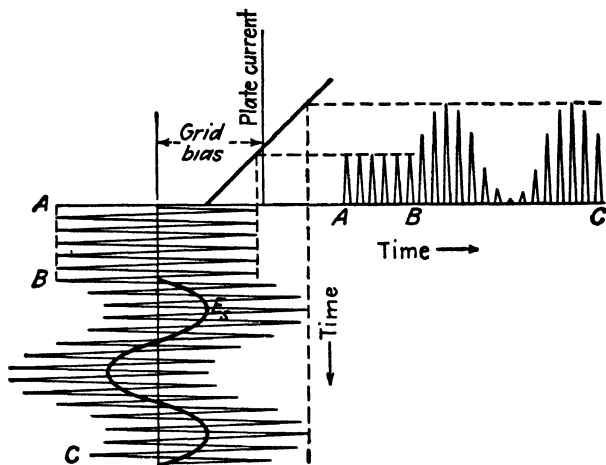



FIG. 227.—Grid-voltage variations and resulting plate-current flow in a system of amplitude modulation by grid injection.

grid must furnish power to the grid circuit. The maximum power requirements can be determined from the analysis of a class C amplifier with the "grid bias" at point (*b*), as previously explained. These power requirements are small—as an estimate, about 10 per cent of the maximum modulated power output. Both amplifiers must be able to deliver this power without distortion, and should have some overload capacity. The direct grid bias E_{cc} must be maintained constant, and it should be from batteries or from a rectifier that has good regulation and maintains the direct voltage constant. The "self-bias" method, using a resistor and parallel condenser (page 285), should not be used. The sideband power does not come from the modulating amplifier as in plate-injection modulation; the sideband power comes from the plate power supply E_{bb}. For this reason, the audio modulating amplifier often is operated in class A. In many instances the modulated

tube may be operated more nearly in class B than class C to reduce distortion. The average plate efficiency is of the order of 30 to 40 per cent, and a given tube will give only about 25 per cent as much modulated output with grid injection as with plate injection.

Amplitude Modulation in Balanced Grid-injection Modulator.—The simplified circuit of a balanced modulator is shown in Fig. 228. This circuit results in lower distortion of the radiated signal than when a single tube is modulated. The method has been used

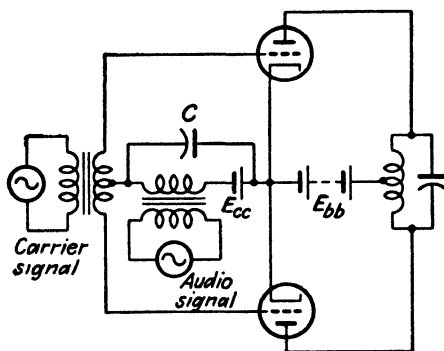



FIG. 228.—Balanced grid-injection amplitude-modulation circuit. Condenser C is a radio-frequency by-pass condenser. The output voltage across the tuned circuit, and the currents in the output coil and condenser will contain the carrier and the two sidebands. If the audio-signal source and the carrier source are interchanged, the output will contain the audio signal and the two sidebands, and the carrier will be suppressed.

in carrier-telephone systems where more than one modulated signal is impressed on, and transmitted along, one pair of telephone wires. This is similar to radio communication in which many signals are transmitted simultaneously through space. The balanced modulator of Fig. 228 has been used in low-power radio transmitters.[1]

As used in radio, the tubes are biased considerably beyond cutoff and the impressed modulating speech or program signal voltage causes the instantaneous grid voltage to vary somewhat, as indicated in Fig. 227. Because the instantaneous grid voltage varies, the output of the tube is as indicated in Fig. 227 and the sidebands are created.

An interesting point regarding the balanced modulator is this:

[1] Kishpaugh, A. W., and R. E. Coram, Low Power Radio Transmitters for Broadcasting, *Proceedings of the Institute of Radio Engineers*, Vol. 21, No. 2, February, 1933.

The carrier voltage drives the two grids 180° out of phase; that is, when one grid is driven less negative, the other is driven more negative, and vice versa. The speech or program modulating voltage, however, drives the two grids in phase. Thus this modulating voltage causes equal changes of current in the identical primaries of the output transformer, and these current changes are in opposite directions. Their magnetic effects cancel as in the push-pull amplifier (page 333), and no components of the modulating frequency exist in the output. In other words, if a speech signal is fed in, as indicated in Fig. 228, the modulated output will contain the carrier and the two sidebands, and the audio component is suppressed. On the other hand, if the two sources of alternating voltage are interchanged and the carrier is injected in series with the grid bias source E_{cc}, then the *carrier is suppressed*, and the output contains the modulating speech or program frequencies and the two sidebands. Thus the circuit of Fig. 228 is useful where single-sideband transmission is desired. The carrier component can be eliminated by the action just described, and the speech and unwanted sideband can be eliminated by filters (page 164).

Amplitude Modulation by Cathode Injection.—In the first system of modulation that was discussed, the carrier was injected between the grid and cathode, and the modulating speech, or program, signal was injected between the plate and cathode. In the second system, both the carrier and modulating signal were injected between the grid and cathode. In the system of cathode-injection modulation, now to be discussed, the carrier is injected between grid and cathode, but the modulating voice or program is injected so that the instantaneous potentials both from grid to cathode and from plate to cathode are varied.

This combined system of grid and plate injection of the modulating signal combines the characteristics of the individual plate-injection and grid-injection methods.[1] The tube is operated in class C, and the driving power and efficiency are midway between the plate-injection and grid-injection systems.

Amplitude Modulation Using Tetrodes and Pentodes.—These tubes (including beam-power tubes) are used extensively as modulators. There are several reasons for this, among them is the fact that the usual neutralization is not necessary.

[1] For further details, see "The Radio Amateur's Handbook," published by The American Radio Relay League.

Tetrodes and Pentodes as Modulators with Grid and Plate Injection.—A typical circuit using a tetrode or beam-power tube is shown in Fig. 229. Condensers C_1 and C_2 are radio-frequency by-pass condensers of the order of 0.001 microfarad. They readily pass the radio-frequency components, but offer high impedance to the modulating speech or program signals. Resistor R is in the circuit to reduce the direct voltage on the screen grid to the proper operating value, which is usually less than the direct voltage needed for the plate. The tuned parallel circuit L-C is the load in the

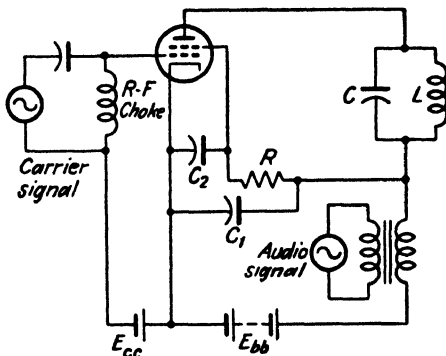



Fig. 229.—A circuit for amplitude modulation using a tetrode or beam-power tube with grid and plate injection. The modulated signal voltage appears across the tuned L-C circuit.

plate circuit. The impedance that this tuned circuit offers to the plate includes the effect of the antenna if one is coupled into it.

The radio-frequency by-pass condenser C_2 prevents the screen-grid potential from varying at the high radio-frequency rate. It will be noted, however, that the screen grid and plate *both* are connected to the source of the modulating speech or program voltage. Thus *both* the screen grid and the plate vary in accordance with these modulating signals. The reason for this is as follows: Over the *usual operating range*, the plate current of a tetrode does not vary much with plate-voltage changes. Thus impressing the modulating signal in series with the plate would be ineffective unless the plate voltage were at a low value, and operation at low plate voltage would cause much undesired distortion because of secondary emission (page 190). But if the modulating signal simultaneously drives both the screen grid and plate, then the plate-current variations will be great; yet the undesired dis-

tortion will not be great. The tube is operated in class C. The power for the sidebands must come from the audio-frequency modulating amplifier.

Tetrodes and Pentodes as Modulators with Grid Injection.—If the carrier wave and the modulating signal speech or program wave are injected simultaneously in the grid circuit, much as in Fig. 226, then modulation will occur and the desired sidebands will be created. For this purpose the tube is operated in class C.

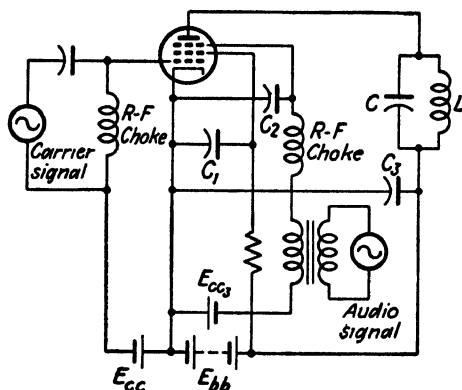



FIG. 230.—A circuit for amplitude modulation using suppressor-grid modulation of a pentode. The modulated signal voltage appears across the *L-C* circuit.

Pentodes as Suppressor-grid Modulators.—A typical circuit arrangement is shown in Fig. 230. As is noted, the screen grid and suppressor grid are by-passed to ground with condensers C_1 and C_2 of the order of 0.001 microfarad, so that the instantaneous potentials of these electrodes cannot vary at a radio-frequency rate. The carrier wave is impressed in the control-grid circuit, and the modulating signal is impressed between the suppressor grid and cathode. Sometimes the suppressor grid is biased, the amount depending on the tube, the circuit, and method of operation. The *L-C* parallel tuned circuit (with the reflected antenna impedance included if an antenna is connected) provides the load in the plate circuit. The radio-frequency choke coils offer high impedance to the high-frequency components.

The tube often is operated in class C, and the action is much like that of the class C triode, with the carrier and the modulating signal both injected into the control grid. In the triode the electron stream constituting the plate current is acted on simul-

taneously by the two signals impressed on the control grid. In the suppressor-grid modulated pentode the effect is much the same, although each wave is impressed on a separate grid. The desired sidebands are created by the combined action of the two grids. If the control and suppressor grids are maintained negative, but little power is required from the carrier and audio sources.

Beat-frequency Oscillators.—These oscillators were not considered in Chap. X because a knowledge of modulation is necessary before their operation can be understood.

In the beat-frequency oscillator, two separate and distinct vacuum-tube oscillators operate at slightly different frequencies,

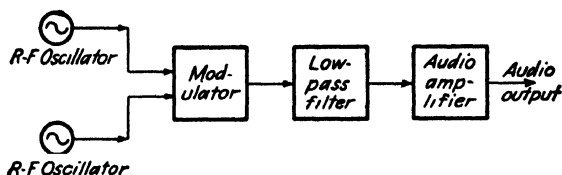



FIG. 231.—Showing the components of a beat-frequency oscillator.

say 100,000 and 101,000 cycles. These two signal frequencies are impressed simultaneously on a vacuum-tube modulator, and sum and difference frequencies, or side frequencies, are created. For the two frequencies just considered, these side frequencies would be 201,000 cycles and 1000 cycles. For an audio-frequency oscillator the 1000-cycle component would be selected by a low-pass filter (page 165), would be amplified, and then would be available for test or other purposes. Of course if the second oscillator were varied slightly so that its frequency was 102,000 cycles, then the output frequency would be 2000 cycles, and so on. A block diagram is shown in Fig. 231.

The beat-frequency oscillator is a very useful device. Since the individual frequencies generated by the two oscillators are high, small coils and condensers may be used. Furthermore, the frequency of only one oscillator need be changed, and this change need be for only a small per cent of the frequency of oscillation in order to cover the entire audio-frequency range. For instance, if the frequency of the variable oscillator is 100,050 cycles, the output of the beat-frequency oscillator will be 50 cycles, and if it is 110,000 cycles, the output will be 10,000 cycles. This simplifies the frequency control; usually it is necessary to vary only the set-

ting of an air condenser in one oscillating circuit. Thus the frequency range can be made *continuously variable*, a very important advantage for some purposes as compared with "stepped" frequency settings.

In considering the beat-frequency oscillator, it again is pointed out that the outputs of the two oscillators must be impressed on a modulator that distorts the waves and creates the desired audio component. Merely "beating" two waves together will not create the desired component (page 418).

Demodulation or Detection.—In a radio system the speech or program signals cannot effectively be transmitted at the frequencies at which they are created. By the process of modulation these signals are translated or moved to higher frequencies. The signals and a carrier wave are impressed on a circuit that causes distortion, and the two sidebands are created. Each of these sidebands by itself contains all the original variations of the message or musical program to be transmitted. In radio broadcasts, the carrier component and the two sidebands are transmitted through space, and are received by the distant radio-receiving set. These signals must be reduced in frequency so that they will be audible when a loudspeaker is driven. This reduction is acomplished by a process called "demodulation," or "detection."

The definitions of demodulation and detection are the same essentially. The term "demodulation" is perhaps more general in meaning, and it is used both in wire telephony and radio. The term "detection" is used extensively in radio. Perhaps demodulation is the better term for the following reason: The information, or program, to be received exists at high frequencies, and must be *reduced* in frequency so that it will be audible. Demodulation implies that the received signals are acted upon by some process. Detection implies that the signal merely is "searched out." In this connection, the point often is overlooked that distortion must be used to create a new and low-frequency band from the radio-frequency signals that are received. Fundamentally, modulation and demodulation are not opposite, but are the same process of nonlinear distortion.

The Demodulation of Amplitude-modulated Waves.—If, at the transmitter, the carrier is *modulated* by a program signal having a band width of from 50 to 10,000 cycles, then the modulated output

will consist of a 1,000,000-cycle carrier component, a lower sideband of from 999,950 to 990,000 cycles, and an upper sideband of from 1,000,050 to 1,010,000 cycles.

When the carrier and the sidebands are received, they are *demodulated* by impressing the three components simultaneously on a circuit that distorts them. In the process of demodulation sum and difference frequencies are created just as in modulation. In modulation, the two frequencies impressed on the modulator are far apart, and the sum and difference frequencies are close together, as listed in the preceding paragraph. In demodulation, the frequencies impressed on the demodulator are close together, and the sum and difference frequencies are far apart.

Thus suppose that a radio-receiving set receives a carrier of 1,000,000 cycles, a lower sideband of from 999,950 to 990,000 cycles, and an upper sideband of from 1,000,050 to 1,010,000 cycles. When these are impressed simultaneously on the demodulator and distorted by it, sum and difference frequencies will be created. The sum frequencies will give bands in the vicinity of 2,000,000 cycles. These are of no use and will be disregarded. The difference frequencies will give two bands that are of importance. The first band will be the difference between the carrier and the lower sideband, 1,000,000 cycles minus the band 999,950 to 990,000 cycles, or 50 to 10,000 cycles. The second band will be the difference between the upper sideband and the carrier, 1,000,050 to 1,010,000 cycles minus 1,000,000 cycles, or 50 to 10,000 cycles. Thus a 50- to 10,000-cycle component has been created from *each* sideband by the process of distortion, and each of these contains the information, or program, that was transmitted in the sideband. The two bands from 50 to 10,000 cycles are in phase and combine to give the audible signal, which is the received message. The audio-frequency amplifier connected between the demodulator is so constructed that it will pass only the desired audible frequencies from 50 to 10,000 cycles, and all other components that exist in the demodulated signal either are rejected or are of little consequence in a well-designed system.

Types of Demodulators for Amplitude-modulated Waves.—Any device that will cause nonlinear distortion and create new frequencies when a signal is impressed on it can be used as a demodulator or detector in radio.

In the early days of radio, many interesting detectors were used.[1] Among these detectors was the crystal. Since this device will rectify (which is a process of distortion), it also will demodulate.

Probably the first use of the thermionic vacuum tube was made by Fleming in 1904, when he used it as a detector of radio-frequency signals. For this purpose the vacuum tube is used in many ways, some of which are obsolete. In modern radio-broadcast receivers the diode detector is used extensively.

A crystal that can be used as a crystal detector in radio (see also Fig. 130, page 225).

The Crystal Detector.—The crystal rectifier was described on page 225. It consists essentially of a suitable crystal, such as one of silicon, and a fine wire that presses firmly on the surface of the crystal. The assembly is mounted in a suitable holder. Modern crystal rectifiers are filled with a wax so that they will not get out of adjustment.

When a *single-frequency* alternating voltage is impressed on the crystal rectifier, the wave is distorted and direct-current and alternating-current components are created. When *two or more* signal waves are impressed simultaneously on the crystal rectifier, nonlinear distortion occurs and side frequencies, or sidebands, are created.

This action is indicated in Fig. 232. If an unmodulated carrier, as from A to B of the figure, is impressed on the crystal, the current through the crystal will be as from A' to B'. If now the input is a modulated wave composed of a carrier and two sidebands, as from B to C, the rectified current through the crystal will be as from B' to C'. By a method similar to that used in Fig. 219, page 408, it can be shown that the rectified wave from B' to C' contains many frequency components, among which is the desired audio-frequency component created by the process of distortion.

Of course, the incoming modulated signal B to C impressed on

[1] A number of very interesting books on the history of radio are available, and these discuss the detectors used the early days of radio. For instance, consult G. G. Blake, "History of Radio Telegraphy and Telephony," Chapman & Hall, Ltd.

the crystal of Fig. 232 had been modulated by a pure sine wave of, perhaps, 1000 cycles. This fact is indicated because the outlines or envelopes of the waves are sinusoidal. Had the modulating signal been speech, or a musical program, then the envelopes would have had the irregular wave shape of the speech or music.

The audio-frequency component of the rectified wave B' to C' of Fig. 232 is the signal desired. It can be separated by a filter or by connecting an audio-frequency amplifier to the crystal output.

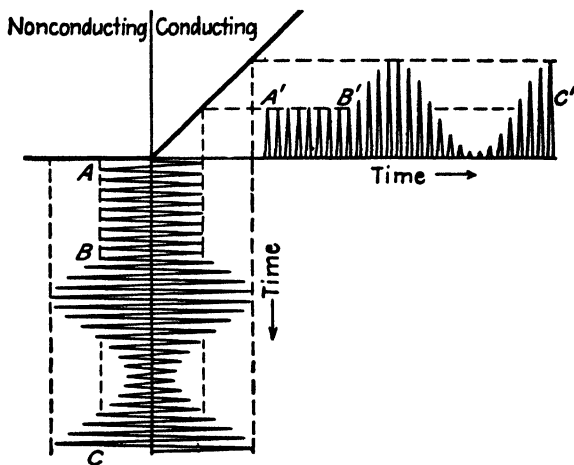



FIG. 232.—When a modulated signal is impressed on a crystal rectifier, current flows for one half cycle, but not the other. This causes distortion, and produces an audio-frequency component. In this figure an ideal crystal rectifier that passes no current in the reverse direction has been assumed.

Crystal detectors are important in modern radio systems which use ultrahigh and superhigh frequencies.

The Diode Detector.—The two-electrode thermionic high-vacuum tube, or diode, is the most satisfactory and the most widely used demodulator, or detector, for ordinary radio purposes. There are many modifications of the basic circuit, a widely used version being given in the simplified circuit of Fig. 233*a*.

The incoming signal consisting of the carrier and the two sidebands is impressed on the diode detector through the tuned radio-frequency transformer at the left.[1] The tube that is widely used is a combined diode and triode.

[1] In the superhetrodyne radio receiver, this would not be the carrier and sidebands at the frequencies at which they would be received (perhaps at about 1,000,000 cycles), but would be the carrier and sidebands at reduced frequency (about 500,000 cycles). This will be explained fully in Chap. XIV .

The diode portion, which is the demodulating, or detecting, part of the circuit, is reproduced in Fig. 233*b*. The signal voltage, composed of the carrier and sidebands, is impressed on the diode portion, and rectification (which is a form of nonlinear distortion) occurs. As a result, sum and difference frequencies are created, the difference frequencies being the original audio frequencies. The shape of the current wave flowing through the rectifier will be

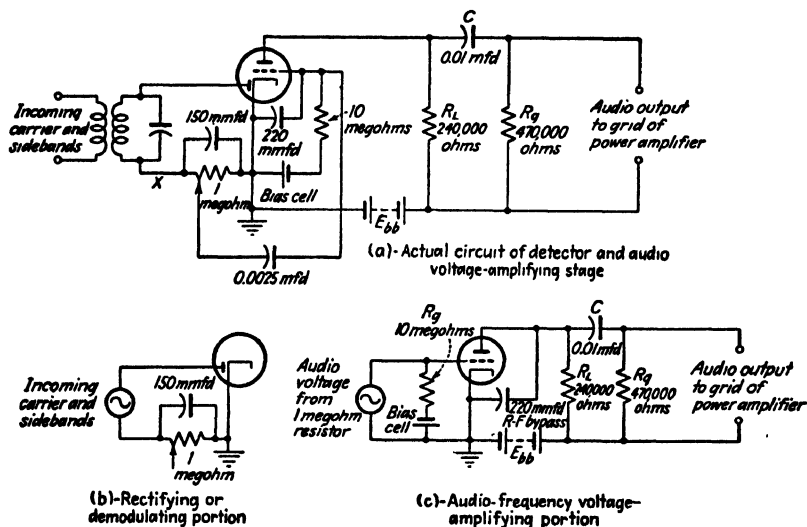



FIG. 233.—Diode detector and audio-frequency voltage amplifier as used in radio-receiving set for amplitude-modulated programs.

much like that of Fig. 232, and it will contain (*a*) a direct-current component, (*b*) an audio-frequency component that is the desired message, or program, signal, and (*c*) radio-frequency components.

All these current components flow through the rectifier and around the circuit of Fig. 233*b*. The reactance of the 150-micro-microfarad condenser is *low* for the radio-frequency components, being about 2000 ohms at 500,000 cycles. Thus the radio-frequency components largely flow through this condenser, and there is negligible radio-frequency voltage drop across it. On the other hand, the reactance of this condenser is *high* for the low-frequency audio components, being about 1,000,000 ohms at 1000 cycles. Thus the audio-frequency components cause, relatively speaking, a large audio-frequency voltage drop across the parallel circuit

composed of the 150-micromicrofarad capacitor and the 1,000,000-ohm resistor.

The circuit of Fig. 233*b* will be recognized as a rectifier with a capacitor-resistor load, and current cutout occurs (page 235). This is at a radio-frequency rate, and for the values of capacitance and resistance used (as given in Fig. 233) causes no difficulty. Of course if a large capacitor were used in parallel with the 1,000,000-ohm resistor, serious distortion of the audio-frequency wave would result. There would be a tendency for the large condenser to charge to the peak of the wave, and to hold the charge because of insufficient time for it to drain off through the 1,000,000-ohm resistor. An alternate viewpoint is that a large condenser would offer low impedance to the audio components, and that negligible audio-frequency voltage would exist across it. The purpose of the 150-micromicrofarad condenser and the 1,000,000-ohm resistor is to separate the desired low-frequency components from the undesired high-frequency components.

As shown on Fig. 233*a*, the 1,000,000-ohm resistor is in reality an adjustable voltage divider. By adjusting the slider, any portion of the audio-frequency voltage drop can be selected and made to appear between the cathode and grid of the triode amplifier portion. If the circuit is redrawn as in Fig. 233*c*, this action is made more evident. The selected audio-frequency component forces a current through the 10-megohm resistor, and the audio-voltage drop across this resistor appears between grid and cathode in series with the **grid-biasing cell.** This special cell is a very small source of direct voltage of about 1 volt, and is incapable of delivering appreciable power.

The triode portion is a high-gain audio-frequency voltage amplifier that often uses a tube having an amplification factor of about 100 and a plate resistance of about 90,000 ohms. The resistor R_L is the load resistor, R_g is the grid resistor, and C is the coupling capacitor to couple the plate circuit of the audio-frequency voltage-amplifier tube to the audio power-output tube. Sometimes audio-frequency transformer coupling is used instead of the R_L, R_g, C circuit.

As previously stated, the diode detector of Fig. 233 is of the type widely used in modern radio-receiving sets. It reproduces the audio-frequency signal with little distortion. The diode section

draws some power from the preceding circuit. The diode detector is particularly useful for the demodulation and detection of radio signals that approach 100 per cent modulation. Some of the other detectors cause very bad distortion of the audio signal when the percentage modulation is high.

Automatic Volume Control.—The diode detector discussed in the preceding section also is a desirable circuit because it is a simple matter to obtain an automatic volume-control voltage from it. This has the following advantages: In the transmission of radio signals from the sending antenna to the receiving antenna, many changes may occur in the transmitting path through space. Some of these changes may occur within a matter of minutes. As a result, the received signals may vary from strong to weak, etc., and this is called **fading** (page 521). Such fading often results in reduced intelligibility and unsatisfactory program reception.

If a circuit could be arranged so that when the received signals were *strong*, the gain of the amplifiers in the radio-receiving set would be *low*, and when the received signals were *weak*, the gain of the amplifiers would be *high*, then a constant audio output of the loudspeaker would result and fading would be counteracted. It would be difficult to use the received sidebands or the demodulated audio-frequency component as a control, because these contain the transmitted signal fluctuations. As has been repeatedly stressed, the carrier-frequency component of the output of a modulator, and of a radio transmitter as well, *does not vary* in amplitude, and this component can be used for control. The strength of the carrier-frequency component is the same before and after modulation in a well-designed amplitude-modulated system.

For automatic volume control a resistor of about 2.0 megohms and a condenser of about 0.05 microfarad are connected *in series* between point x of Fig. 233*a* and ground, which is in reality from point x to the cathode. This gives the equivalent circuit of Fig. 234*a*. Assume that the radio-receiving set, of which Fig. 234*a* is the detector, is tuned to a distant radio-transmitting station. Also assume that for the moment no modulating audio signal is being impressed on the transmitter. The distant transmitter will then be sending out an unmodulated single-frequency carrier wave. This wave will be received, and will be rectified by the diode. As a result, a direct-current component, *the magnitude of which is proportional to the received carrier*, will flow as indicated by the

arrow in Fig. 234*a*. This direct current will cause a voltage drop having the polarity shown.

Now assume that the strength of the received *unmodulated* carrier varies relatively slowly in magnitude because of variations in the transmitting path through the space between the radio transmitter and receiver. Over the interval of time that the received carrier is strong, a large rectified current will flow, and the left end of the 1.0-megohm resistor will be quite negative. Over the interval of time that the received carrier is weak, the rectified

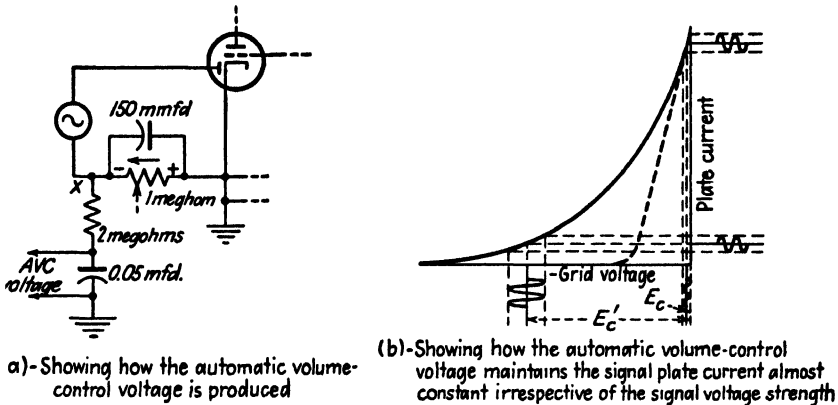



FIG. 234.—Circuit showing how the automatic volume-control voltage is obtained and how it is used.

current will not be so strong, the voltage drop across the 1.0-megohm resistor will be less, and the left end of the 1.0-megohm resistor will be less negative.

The 0.05-microfarad condenser is connected to this negative potential through a 2.0-megohm resistor. The capacitance and series resistance are so large that only relatively slow variations in the charge on the condenser electrodes are possible. This means also that only slow variations in voltage across the condenser can occur. When the received carrier is strong, and the voltage drop across the 1.0-megohm resistor is large, the voltage across the 0.05-microfarad condenser tends to build up to this large voltage. When the received carrier is weak, and the voltage drop is low, the potential of the condenser tends to fall. The capacitance of this condenser, and the resistance of the 2.0-megohm resistor in series with it, are so selected that the voltage change across the 0.05-microfarad condenser *cannot follow* the audio-frequency voltage

changes that occur across the 1.0-megohm resistor *when a modulated signal is being received*, as explained in the preceding section.

The voltage across the 0.05-microfarad condenser, which varies in accordance with the relatively slow changes in the received signal, is used to control the amplification in a radio-frequency amplifier preceding the detector stage. This is done by using this voltage, called the **automatic volume-control (avc) voltage,** to vary the grid bias on a special amplifier tube.

Conventional voltage-amplifying triodes, tetrodes, and pentodes have what is called **sharp cutoff characteristics,** as shown dotted in Fig. 234*b*. That is, the control grid voltage–plate current characteristic (page 192) is relatively straight and falls rapidly to zero as the control grid is made more and more negative. By contrast, the corresponding characteristic of the special tube on which the avc voltage is impressed falls gradually to zero, as indicated by the full line of Fig. 234*b*.

As indicated in this figure, no matter where the conventional tube is biased, a given grid-voltage change causes about the same plate-current change except in the vicinity of cutoff. The amplification is almost constant as Fig. 102, page 187, indicates. But with the special tube, a weak grid signal voltage with a bias E_c causes essentially the same plate-current change as a strong grid signal voltage at a bias E_c'. This special tube has a variable amplification factor, the value depending on the grid bias. This tube is called by several names, such as the **variable-mu tube,** the **supercontrol tube,** and the **remote cutoff tube.** This variable amplification factor is obtained by using a special control-grid construction.

Thus when the received radio signal is strong, the bias is large and but little amplification results. When the received radio signal is weak, the bias is small and much amplification results. By this action the volume changes caused by fading of the received radio signal are compensated for automatically. On the other hand, holding the volume constant does not offset the distortion that sometimes accompanies fading. Manual volume control is accomplished by adjusting the position of the slider on the 1.0-megohm volume control of Fig. 233*a*.

The Triode Detector.—Demodulation with triodes can be accomplished by either the grid current or the plate current. If the grid current is used, then demodulation is caused by the nonlinear

grid current–grid voltage characteristics. If the plate current is used for demodulation, then it is accomplished by the nonlinear grid voltage–plate current characteristic. In considering the circuits used, it should be remembered that demodulation and modulation are the same fundamental process.

Grid-current Demodulation.—This is the so-called **grid leak–condenser detector** so extensively used in early radio receivers. The tube usually is initially unbiased. When an *unmodulated* signal is impressed on the tube of Fig. 235*b*, this carrier wave is rectified and the resulting direct current flows through the grid resistor R_g. This direct current will cause an IR drop, as indicated on the circuit diagram, fixing the operating point at P.

The grid voltage–grid current curve is as shown in Fig. 235*b*. As a result of the curvature, the grid current that flows is distorted as indicated. As a result of the distortion, sum and difference frequencies result, and the desired audio-frequency component is created. The condenser C_g is of such value that it offers negligible reactance to the radio-frequency components, but high reactance to the audio-frequency components. The audio-frequency components must, therefore, flow through the grid resistor and cause an audio-frequency voltage drop across it. This audio voltage is in reality between the grid and cathode, and is *amplified* in the plate circuit of the tube. The small condenser in the plate circuit by-passes all radio-frequency components, and the desired amplified audio output voltage is available across R_L for further amplification. This method is not satisfactory for waves with a high degree of modulation, because excessive audio-frequency distortion results. This method is also called **weak-signal grid detection** and **square-law grid detection.** This last term is used because of the particular mathematical equation for the grid voltage–plate current curve. For this type of detection R_g is from 1 to 5 megohms, and C_g is from 100 to 250 micromicrofarads. Tetrodes or pentodes may be used.

As emphasized in the preceding discussion, a very weak modulated signal wave is impressed. If, however, a relatively strong modulated signal wave is impressed, then operation is as shown in Fig. 235*c*. Here it is assumed that the curvature is of no consequence. With a strong-signal operation, the grid circuit acts like a diode rectifier. The action of the grid resistor R_g and the grid condenser C_g is fundamentally the same. Because the created

Fig. 235.—Circuit and diagrams for grid-current demodulation or detection.

audio-frequency component flows through the grid resistor, the instantaneous grid voltage varies in accordance with the audio signal, and amplification occurs in the plate circuit. For this type of detection R_g is from 100,000 to 500,000 ohms, and C_g is from 50 to 100 micromicrofarads. This is called **strong-signal detection, power detection, and linear detection.** This last term may be confusing, because it has been explained that both modulation and demodulation are forms of nonlinear distortion. The term "linear detection" is applied because the curve of Fig. 235*c* may be considered straight when strong signals are applied. But it is straight to *one-half* the wave only, and there is nonlinear distortion as required to create the desired audio component. Usually less audio distortion results when strong-signal detection instead of weak-signal detection is used. Both these systems draw radio-frequency power because grid current flows.

Plate-current Demodulation.—The simplified diagram of a plate-current detector is shown in Fig. 236*a*. The grid is biased negatively, so that negligible grid current flows and negligible power is drawn from the preceding stage. The distortion required to produce the audio component from the impressed carrier and sidebands results from the nonlinear grid voltage–plate current characteristics.

If the circuit is designed to handle relatively weak amplitude-modulated signals, operation is as in Fig. 236*b*. Because of the curvature of the grid voltage–plate current characteristic, the output current is distorted, as shown, and will contain the desired audio component. The radio-frequency components in the plate current will be by-passed by condenser C, which is of the order of 1000 micromicrofarads. The demodulated audio component will pass through R_L, which is of the order of 50,000 ohms. This resistor is for resistance coupling the detector into the grid of the following audio amplifier. Of course audio-frequency transformer-coupling may be used. This is called **weak-signal plate detection** and **square-law plate detection.**

If the conditions of operation are so arranged that the signal to be demodulated is impressed as in Fig. 236*c*, and if the signal is relatively strong, then the plate current will be as shown. The desired audio signal is created from the carrier and sidebands because of the distortion caused by plate current flowing on one half cycle but not on the other. The radio-frequency components are

FIG. 236.—Circuit and diagrams for plate-current demodulation or detection.

by-passed by condenser C of Fig. 236, and the audio signal current flows through the load resistor R_L, where it causes a voltage drop that is amplified and then used to drive the power stage and the loudspeaker. This method is called **strong-signal plate detection** or **linear plate detection.**

Frequency Modulation.—As explained early in this chapter, if the angle ϕ of Eq. (121) is varied, then angle or angular modulation results. Frequency modulation and phase modulation are very closely related types of angle modulation. Of these two, frequency

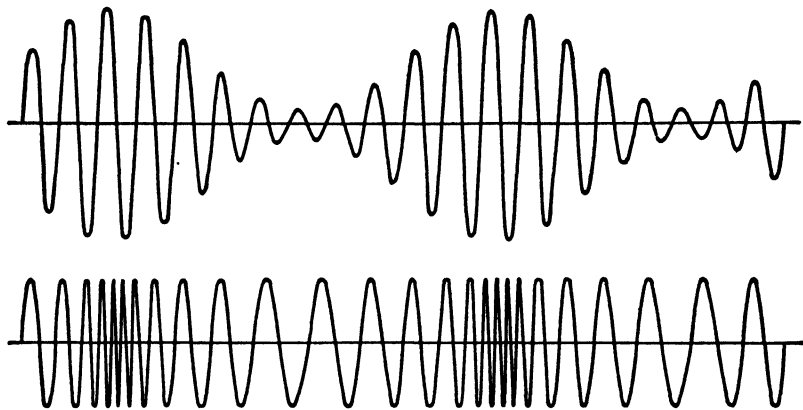



FIG. 237.—Showing an amplitude-modulated wave (above) and a frequency-modulated wave (below). In each instance the modulating signal is a pure sine wave such as 1000 cycles.

modulation is the more widely used, and accordingly will be considered first, and in more detail.

An amplitude-modulated wave and a frequency-modulated wave are shown in Fig. 237. Both these waves have been modulated with a sinusoidal single-frequency wave, such as a 1000-cycle signal.

In frequency modulation the output from the modulator and the signal radiated at the radio transmitter are pure sine waves if no modulating signal is impressed. In other words, if the transmitter is unmodulated, the output is the carrier frequency, just as in amplitude modulation. When a frequency-modulation transmitter is modulated with a pure sine wave (for example, by whistling into the microphone), then the output appears as in Fig. 237.

At each instant the frequency is different from that of the preceding instant in a frequency-modulated wave. The instantaneous

frequency makes "excursions," or "swings," from the center carrier frequency that is assigned to the particular transmitter. For instance, the assigned basic carrier frequency may be 100,000,000 cycles (100 megacycles), and the maximum swing or shift may be 75,000 cycles each side of this value. The *magnitude* of the frequency shift depends on the *magnitude* of the modulating speech, or program, signal. The *number* of complete swings per second (that is, the *frequency* of the swings) depends on the *frequency* of the modulating signal. Thus a frequency-modulated transmitter might be adjusted so that for one cycle of the loudest modulating signal the frequency would change from 100,000,000 cycles to 100,075,000 cycles, to 99,925,000 cycles, and back to 100,000,000 cycles. For a modulating signal of 1000 cycles, one thousand of these complete shifts would occur per second. Of course if a complex speech or program wave were used to modulate the transmitter, then the maximum frequency changes, and the rates at which these occurred, would depend on the nature of the complex modulating signal.

The Nature of a Frequency-modulated Wave.—As was mentioned on page 406, it is sometimes considered that the amplitude-modulated wave of Fig. 237 is merely the carrier-frequency wave that varies in amplitude. But as explained, the component of assigned carrier frequency does not vary with modulation; the sidebands vary in accordance with the modulating signal and convey the information or musical program. In much the same way, it is sometimes considered that the frequency-modulated wave is a rather simple wave that merely varies in frequency. But the frequency-modulated wave is far from simple when it is studied in detail.

An analysis shows that if the modulating signal is a sine wave, then the frequency-modulated wave of Fig. 237 consists of a carrier component and side frequencies. In this respect, a frequency-modulated wave is similar to an amplitude-modulated wave, but this similarity does not extend far. In amplitude modulation, the magnitude of the component of carrier frequency remains constant before and after modulation. In frequency modulation, the magnitude of the component of assigned carrier frequency varies with modulation (this is in contrast with the popular viewpoint, which considers that the frequency-modulated wave is a "carrier" varying in frequency only). In amplitude modulation

by a pure sine wave, *two* side frequencies are created; these are sidebands when speech, music, or other signals are used to modulate the transmitter. In frequency modulation (theoretically) an infinite number of single side frequencies is created when the modulating signal is a pure sine wave. The question immediately arises: Why is not an infinitely wide band covered by a frequency-modulated station if an infinite number of side frequencies is created? The answer is that this is theoretical only. A frequency-modulation transmitter must be kept within the band of frequencies assigned to it, and it could not be permitted to radiate an infinitely wide band.

It is interesting to note that the original work on frequency modulation was done to ascertain if a frequency-modulation channel could be narrower than an amplitude-modulation channel. It can be shown both mathematically and experimentally that for satisfactory operation a frequency-modulation transmitter must have a wider frequency band in which to work than is needed by a comparable amplitude-modulated transmitter.

If the modulating signal is speech, music, or code, then the frequency-modulated wave contains sidebands instead of sinusoidal side frequencies. An analysis of these sidebands is somewhat involved, and it will not be attempted. The point of importance is that the total band width required by a frequency-modulated signal is essentially the same whether this band width is explored by using single audio-frequency components for modulation, or whether a complex audio speech or music wave is used.[1]

As was mentioned on page 440, a typical frequency-modulated wave may vary to a maximum of 75,000 cycles above, or to a minimum of 75,000 cycles below, the mid, or carrier, frequency assigned to the station. The maximum change in frequency from the mid or unmodulated carrier frequency is the **frequency deviation,** sometimes designed f_d. This limit is determined by design and adjustment. For instance, frequency-modulation broadcast transmitters are designed and operated so that the peak value of the strongest modulating signal drives the instantaneous frequency to plus and minus 75,000 cycles with respect to the basic carrier frequency. This is, however, merely a limit set by standard practice; the equipment *could* be designed so that the frequency

[1] Those interested in additional details of frequency and phase modulation should consult the reference listed as a footnote on p. 406.

deviation is 50,000 cycles, or any other value. In fact, narrower limits are used in police radio systems, etc.

If a single-frequency modulating wave is used, the *number* of the side frequencies that are of appreciable magnitude, and the *magnitude* of these side frequencies, are determined by the **frequency modulation index,** sometimes designated m_f. This is the ratio of the maximum frequency deviation f_d to the frequency f_m of the modulating wave, or

$$m_f = \frac{f_d}{f_m}. \tag{123}$$

For instance, if the maximum frequency deviation f_d is 75,000 cycles, and the frequency of the sinusoidal modulating signal f_m is

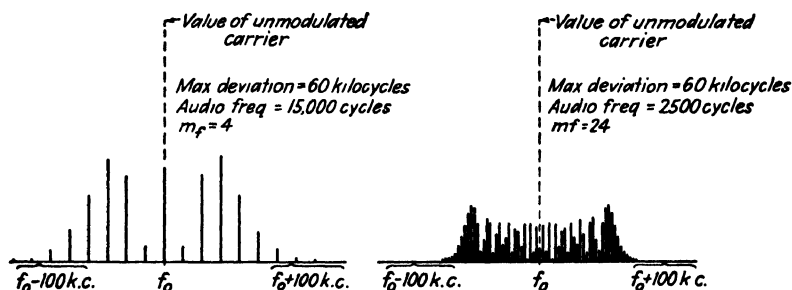



FIG. 238.—A frequency analysis of two frequency-modulated signals. These show that a frequency-modulated wave contains many components and that the magnitude of the carrier component changes. (*From reference listed on page 406, courtesy of Electrical Engineering.*)

10,000 cycles, the frequency modulation index is 75,000/10,000 = 7.5. Apparently the letter symbols have not been standardized. The symbols m_f, f_d, and f_m are used sometimes, and have been adopted here as reasonable designations.

If the maximum deviation of a certain frequency-modulation radio transmitter is a given value, then the frequency spectrum will depend on the frequency of the modulating signal. That is, it depends on the modulation index as given by Eq. (123). Two frequency spectrums are shown in Fig. 238. These are reproduced from the article referred to in the footnote on page 406. In these the maximum frequency deviation is 60,000 cycles. The first is for a sinusoidal modulating frequency of 15,000 cycles and a modulation index of 60,000/15,000 = 4. The second is for a modulating frequency of 2500 cycles and a modulation index of 60,000/2500 = 24. Several important characteristics are: (*a*) that the magnitude of the carrier-frequency component is different for the different

modulation indexes; (*b*) that the number, magnitudes, and frequencies of the various side frequencies also are different; (*c*) that the higher the frequency of the modulating signal, the less the number of side frequencies, and yet the farther away from the carrier are appreciable components found; (*d*) that although the maximum frequency deviation for these figures is 60,000 cycles, the bandwidth required for the complete frequency-modulated wave would be about 200,000 cycles.

Although the graphs of Fig. 238 are for a maximum deviation of 60,000 cycles, they indicate the fundamental nature of the frequency-modulated waves that would result if the maximum deviation were 75,000 cycles, as it is in modern broadcast frequency-modulation transmitters. When this deviation is used, the band width required is about 200,000 cycles for each frequency-modulation station. The audio-frequency modulating signal used in modern broadcast frequency modulation is from about 30 to 15,000 cycles.

Methods of Frequency Modulation.—Although the practical application of frequency modulation is new as compared to amplitude modulation, many systems of frequency modulation have been developed. An analysis of each of these would require a lengthy discussion. Hence, the basic principles common to the various systems will be presented. Some of the means of obtaining a frequency-modulated wave may be considered as *indirect*. The desired wave is synthesized, or built up step by step. Other methods may be considered as *direct* frequency modulation.

Indirect Frequency Modulation by Wave Synthesis.—It was shown by Roder,[1] in 1931, that a frequency-modulated wave would result (*a*) if the modulating audio-frequency signal were so distorted that the magnitude of each component was inversely proportional to the frequency; (*b*) if this were used to produce amplitude modulation and create two sidebands; (*c*) if the two sidebands were shifted 90° with respect to the carrier.

A system of frequency modulation working on this principle was announced by Armstrong[2] in 1936. A simplified block diagram of the method is shown in Fig. 239. The "carrier" frequency of

[1] Roder, Hans, Amplitude, Phase, and Frequency Modulation, *Proceedings of the Institute of Radio Engineers*, Vol. 19, No. 12, December, 1931.

[2] Armstrong, E. H., A Method of Reducing Disturbances in Radio Signaling by a System of Frequency Modulation, *Proceedings of the Institute of Radio Engineers*, Vol. 24, No. 5, May, 1936.

several hundred thousand cycles is impressed on the balanced *amplitude* modulator. The incoming amplified audio signal passes through the frequency-distorting network that makes the amplitude of the various frequency components inversely proportional to the frequency. This is then impressed on the balanced *amplitude* modulator, which suppresses the carrier and gives out two sidebands (page 421).

The two sidebands are then shifted 90° in phase and are combined with the carrier that came directly from the crystal oscillator. The result is a frequency-modulated wave that varies a small

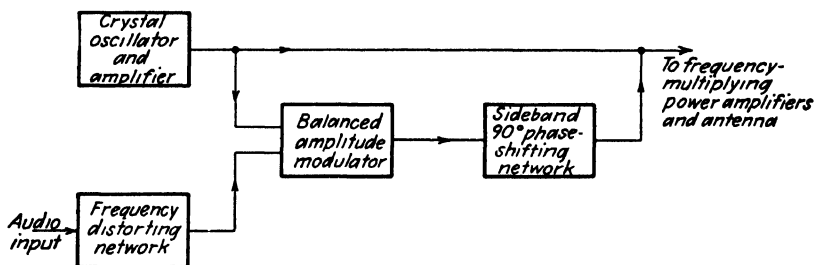



FIG. 239.—A system of frequency modulation by which the desired wave is built up or synthesized.

amount about the basic carrier, usually of several hundred thousand cycles. This wave then is impressed on power frequency multipliers (page 349) that do several things: (*a*) increase the power in the wave; (*b*) increase the basic frequency until the desired carrier frequency is reached; (*c*) increase the frequency modulation index so that a wave centered on, perhaps, 100,000,000 cycles and with a frequency deviation of 75,000 cycles is created. Modifications of this method are used in modern broadcast transmitters.

Direct Frequency Modulation Using the Reactance Tube.—The reactance tube is a conventional vacuum tube so arranged in a circuit that it produces a leading, or lagging, current and hence in its electrical effects is equivalent to a coil or condenser. The reactance can be made to vary in accordance with the magnitude of the modulating audio signal. If this voice-controlled variable reactance is connected across the tuned circuit that determines the frequency of an oscillator (Fig. 240), then the output of the oscillator can be made to vary in accordance with the voice, and will be frequency-modulated.

A simplified circuit of a reactance-tube frequency modulator is shown in Fig. 240*b*. The parallel circuit *L-C* determines the

frequency of the oscillator. A series combination of condenser C_1 and resistor R_1 is connected across the tuned L-C circuit. The capacitance is very low and the C_1-R_1 combination is selected to be a high value of *almost pure capacitive reactance.*

When the tube at the right is oscillating, there will be a voltage at the frequency of oscillation across the tuned circuit L-C. Since the C_1-R_1 circuit is highly reactive and is in parallel with the L-C oscillating circuit, a very small current, leading the voltage across

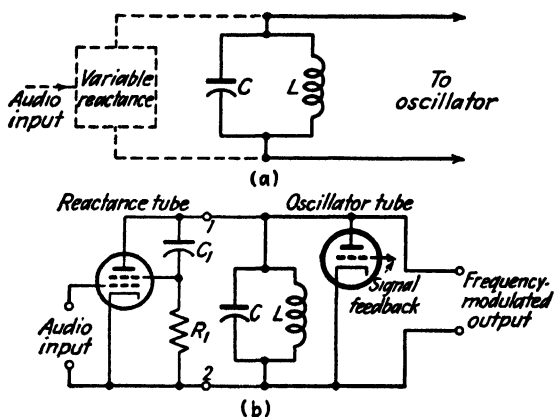



FIG. 240.—The basic principle of reactance-tube frequency modulation (above) and a simplified circuit (below).

points 1–2 by almost 90°, will flow through C_1-R_1. The voltage drop across R_1 will be *in phase* with the current through C_1-R_1. This voltage drop will appear between one of the control grids and the cathode of the reactance tube. A tube, such as the 6L7, with *two* control grids is used. The tube has been simplified in Fig. 240*b*.

The voltage drop across R_1 that is impressed on the tube will be amplified and will cause an alternating plate current to flow toward points 1–2. This current will *lead* the voltage across points 1–2 and across the tuned L-C circuit by essentially 90°. This is because from Fig. 157, page 274, the plate current of a tube is *in phase* with the grid voltage when the tube works into a resistance load, and the tuned L-C circuit is a resistance load at the resonant frequency. The reactance tube and its circuit, as connected to the left of points 1–2, act therefore like a capacitor. Because of this, the reactance tube can influence the frequency of

oscillation of the **master oscillator tube** at the right. The presence of the reactance tube causes the same effect as if capacitor C were increased.

The speech or program signal voltage that is to frequency-modulate the transmitter is impressed as indicated between the other control grid and cathode of the reactance tube. The variations in magnitude of the audio signal cause variations in the magnitude of the leading plate current and, hence, the equivalent capacitance across points 1–2 of Fig. 240*b*. In this way the magnitude of the instantaneous audio signal causes corresponding changes in the magnitude of the instantaneous frequency. Also, these changes in frequency occur at the audio rate, and hence the tube at the right produces a frequency-modulated radio signal.

If the positions of the resistor R_1 and capacitor C_1 of Fig. 240*b* are interchanged, and if R_1 has high resistance and C_1 causes relatively low capacitive reactance, then the equivalent reactance connected across points 1–2 will be inductive, and the oscillator will be frequency-modulated because of an equivalent variable inductance across its tuned circuit.

Reactance-tube frequency modulators, and somewhat similar resistance-tube frequency modulators, are used in several modern frequency-modulation transmitters. Sometimes a combination of two reactance tubes is used.

Direct Frequency Modulation Using Special Tubes.—Several methods of frequency modulation have been developed, which use special electronic devices. One method that has been adopted uses a special tube called the **Phasitron.** The action of this tube will be described now.[1]

Electrons are emitted from a central portion of an indirectly heated cathode that extends up through special focusing electrodes, anodes, and a neutral plane (Fig. 241). The electrons as a group flow out from the central portion of the cathode in the form of a *thin electron disk* that extends outward in all directions in a plane at right angles to the cathode. The electrons are collected by cylindrical anodes about the cathode.

Surrounding the cathode and below the disk of electrons are

[1] It is difficult to present a complete explanation of the Phasitron in a few paragraphs. Those interested in a detailed discussion should consult publications of the General Electric Company covering the subject. See also Robert Adler, A New System of Frequency Modulation, *Proceedings of the Institute of Radio Engineers*, Vol. 35, No. 1, January, 1947.

special deflector electrodes composed of wires. A positive potential on these wire deflector electrodes will attract the electron disk and bend the edges down. A negative potential on these electrodes will repel the electron disk and bend the edges up.

But the deflector electrode wires are not all made positive and then all made negative. Instead, every third deflector wire is

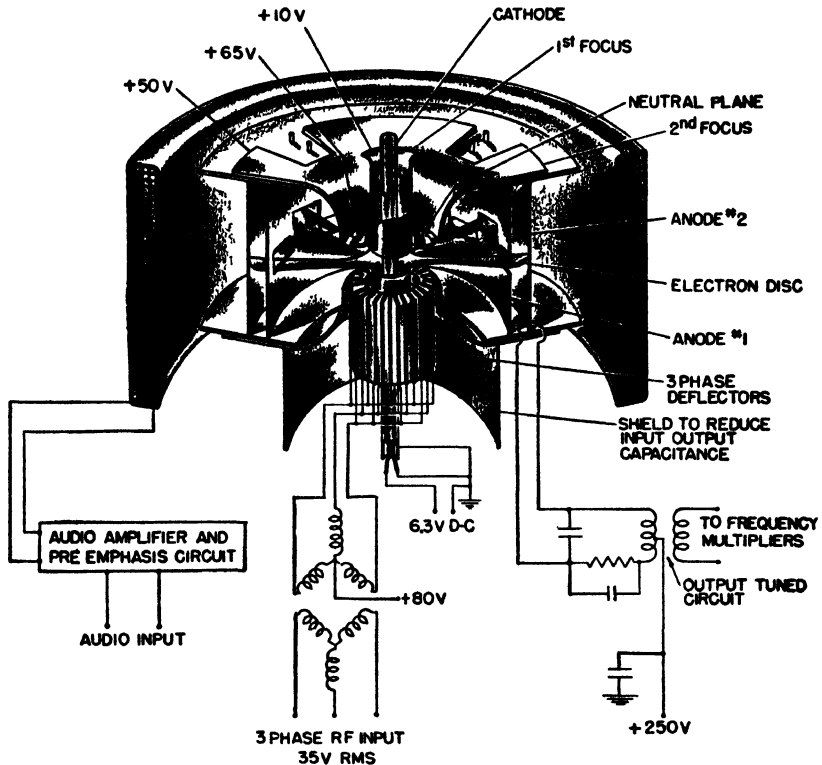



FIG. 241.—Schematic diagram of the Phasitron used in frequency modulation, and the simplified circuit connections. (*General Electric Co.*)

connected to one phase of a three-phase radio-frequency voltage. This voltage is obtained from special networks that are connected to a crystal oscillator. This crystal-controlled three-phase voltage when applied to the deflector wires does not cause the entire edge of the electron disk to move up or down. Instead, the three-phase voltage causes *portions* of the edge of the electron disk to move up, and at the same time causes portions of the edge to move down. This is because the separate voltages of a three-phase system are

120° out of phase, and some wires will be negative when others are positive.

A three-phase voltage when applied to the deflector wires causes another and *very important* effect. The polarities of the wires not only vary from positive to negative but also, the polarities rotate or travel around the wires. In other words, a *rotating* alternating electric field is produced under the electron disk.

The positive electron-collecting anodes 1 and 2 will be considered now. If no voltage is applied to the deflecting wires below the electrons, then the electrons flow out in a flat disk from the

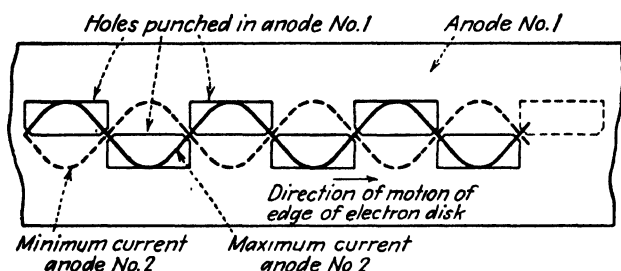



FIG. 242.—As the electron disk with the scalloped edge of the Phasitron moves along, the current flows first to one anode and then to the other. Two positions of the moving electron disk are shown by sine waves.

cathode to the positive anode. However, anode 1 has rectangular holes in it. These holes are placed around anode 1 alternately above and below the region where the electrons would strike if they were in a flat disk, as they would be with no three-phase deflecting voltage applied. Anode 2 is a metal cylinder around anode 1, and anode 2 has no holes in it.

With the three-phase voltage applied to the deflector wires, the electrons are not in a flat disk, but are in a disk with sinusoidal "scollops" around the edge, and these sinusoidal scallops move around the disk at a high-frequency rate determined by the crystal oscillator frequency.

In Fig. 242, anode 1 is shown "flattened out," and anode 2 is assumed to be immediately behind it. The holes in anode 1 are shown. Electrons that go through these holes are collected by anode 2. With the three-phase alternating voltage applied to the deflector wires, the scalloped edge of the electron disk moves along as indicated. As the edge moves from the dotted position, to the position shown by the solid line on Fig. 242, and on to the dotted

position, etc., the sinusoidally scalloped edge of the electron disk causes the current *between the two anodes* to vary sinusoidally. These two anodes are connected to a parallel tuned circuit, which performs in the conventional manner.

The method of modulating will be explained now. This is done by amplifying the program signal, distorting the signal (page 454) so that the magnitude of each component is inversely proportional to the frequency, and then passing the signal current through the coil that surrounds the tube. When the modulating signal current is zero, a given electron will travel a given path to the anodes. When the modulating signal current is in a given direction, the path of the electrons will be curved (page 395), and they will strike ahead of the place they would hit with no modulating signal. When the modulating signal current is in the opposite direction, the electrons will strike behind the place they would strike with no signal. Of course the magnitudes of the shifts depend on the magnitude of the modulating signal, and occur at the same frequency. These shifts, in accordance with the modulating signal, modulated the output of the tuned anode circuit. This is connected to power frequency multipliers that increase the basic frequency to about 100,000,000 cycles, increase the maximum frequency deviation to 75,000 cycles, and provide the necessary power.

The Demodulation of Frequency-modulated Waves.—The demodulator or detector in a radio receiver for frequency-modulated signals must, of course, so act on the received modulated waves that the original audio-frequency modulating speech, or program, is obtained. There are several ways of accomplishing this, but only the most widely used methods will be discussed.

The circuit is shown in simplified form in Fig. 243*a*. This demodulating circuit is called a **discriminator.** The primary and secondary of the transformer are tuned to the same frequency, which is the midfrequency, or frequency of the *unmodulated* wave.

Received Signal at Midfrequency.—The first analysis will be made at the instant the received signal is at the midfrequency. The voltage E_p across the primary will cause a current I_p in the primary coil, and this will lag essentially 90° behind the voltage E_p. This primary current I_p will induce a voltage E_i in the secondary coil. This induced voltage E_i will lag the primary current I_p by 90°. Since the secondary coil and condenser are tuned to resonance,

and since the induced voltage is in series in the secondary coil, a current I_s in phase with E_i will flow around the secondary. This will cause a voltage drop across the condenser, this voltage drop will lag the current I_s by 90°, and this voltage is the voltage E_s across the secondary tuned circuit. The vector positions are as shown in Fig. 243*b* in the drawing at the left.

Received Signal below the Midfrequency.—When the instantaneous frequency of the frequency-modulated wave is *below* that of the carrier (midfrequency), the following analysis holds. The voltage E_p will cause a current I_p to flow, and this will induce a voltage E_i in series in the secondary coil. But now the received frequency is below the frequency to which the secondary is tuned, and an excess of capacitive reactance will exist in the secondary. Because of this, the voltage E_i induced in series in the secondary coil will cause a *leading* current I_s to flow. This will cause a voltage drop E_s across the condenser, as shown in the center diagram of Fig. 243*b*.

Received Signal above the Midfrequency.—At the instant that this occurs, there will be an excess of inductive reactance in the tuned secondary, I_s will lag E_i, and the secondary voltage will be displaced, as shown by the diagram at the right of Fig. 243*b*.

The Demodulation Process.—It has been shown that, as the received frequency-modulated signal falls below and rises above the midfrequency value, the voltage E_s across the secondary is shifted in phase with respect to the voltage E_p across the primary. It now remains to be explained how these variations are converted into the original audio-frequency modulating signal.

The two portions of the circuit of Fig. 243*a* are grounded at the bottom of the primary coil and at the bottom of resistor R_1 (grounds not shown); also, the center of the secondary coil is connected to the top of the primary coil, as shown, through a condenser of negligible reactance at the frequencies involved. This places the vector sum of the primary voltage E_p and one-half of the secondary voltage E_s between *each diode plate and cathode.* These voltages have been designated E_1 and E_2 for the two diode plates.

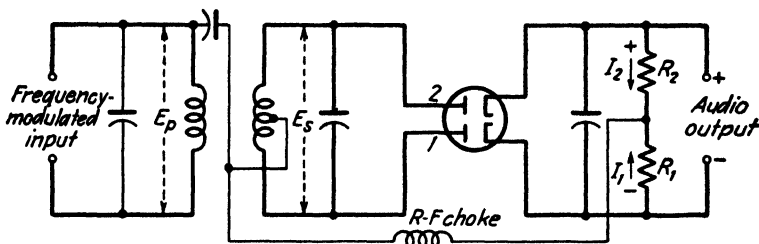
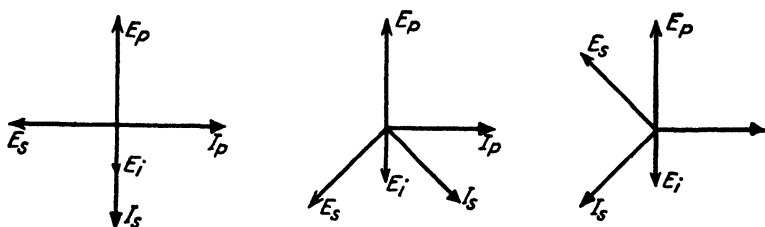
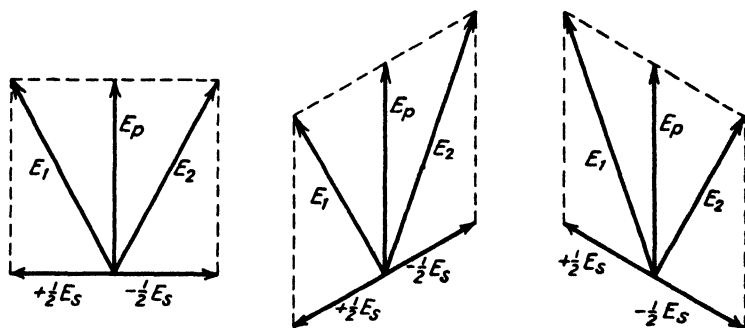
When the received signal is *at the midfrequency*, or carrier, value, conditions are as shown in the diagram at the left of Fig. 243*c*. Both diode plates have alternating voltages of the same magnitude E_1 and E_2, and the rectified currents that flow through the two halves of the cathode resistor will be equal and opposite. Hence, no voltage will exist between the output terminals. This is just

(a)-Simplified circuit of discriminator

(b)-Vector diagrams of currents and voltages in primary and secondary

(c)-Showing the voltages E_1 and E_2 across the diode elements as the frequency-modulated signal frequency varies

FIG. 243.—Circuit and vector diagrams for a discriminator used to demodulate frequency-modulation signals. The current directions and voltage polarities in (a) are for the condition that the instantaneous frequency is below the midfrequency. In the actual circuits the lower end of the primary coil and the lower end of R_1 are connected together by grounding or otherwise. (See the circuit of the ratio detector.)

what is wanted. When the received signal is on the midfrequency, or carrier, it is because at that instant the audio modulating signal is zero; hence, if the original modulating wave is to be recreated by the discriminator circuit of Fig. 243*a*, the output should be zero at the instant under consideration. The combined rectified direct current flows through the radio-frequency choke to the center of the secondary winding, and to the diode plates. The condenser in the cathode circuit by-passes high-frequency components.

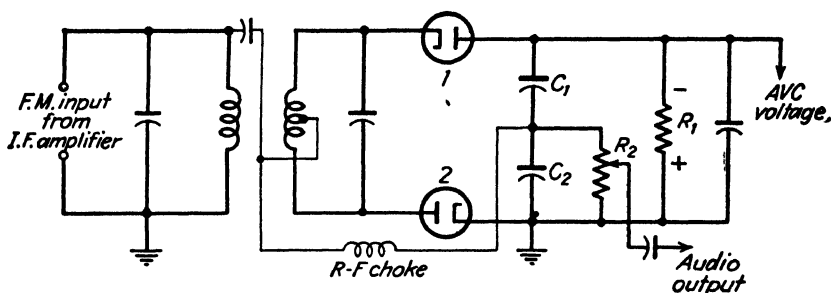
When the received signal *falls below the midfrequency* value, conditions are as in the center diagram of Fig. 243*c*. Note that now E_2 exceeds E_1. If connections are such that E_2 is between the plate and cathode of diode 2, then current I_2 will exceed current I_1, and since the two resistors R_2 and R_1 are identical, there will be a greater voltage drop across R_2 than R_1. Hence, the upper end of R_2 will be more positive than the lower end of R_1. This causes a net polarity, as indicated by the + and − signs across the audio-frequency output terminals.

When the received signal *rises above the midfrequency value*, conditions are as at the right in Fig. 243*c*. Now voltage E_1 exceeds E_2, and current I_1 exceeds I_2. For these conditions the lower end of resistor R_1 will be more positive than the upper end of R_2, and the polarity across the audio-frequency output terminals will be reversed.

In a frequency-modulated wave the magnitude of the instantaneous frequency deviation from the midfrequency depends on the magnitude of the modulating speech or program signal. In Fig. 243*c*, the magnitude of the net output voltage across the R_1-R_2 combination depends on the magnitude of the frequency deviation. In a frequency-modulated wave, the number of frequency excursions, or deviations, per second depends on the frequency of the modulating speech, or program, signal. In Fig. 243*c*, the rate at which the conditions change, as shown by the vector diagrams, and the rate at which the net output voltage changes depends on the number of frequency excursions, or deviations, of the frequency-modulated signal. Thus the discriminator circuit of Fig. 243*a* directly reproduces the desired audio speech or program signal from the received frequency-modulated wave. In a frequency-modulation radio-receiving set the discriminator is preceded by a limiter to be discussed in Chap. XIV.

A circuit sometimes used for demodulating frequency-modulated

signals is the **ratio detector** shown in an accompanying diagram. This circuit has features in common with Fig. 243*a*, and much of the preceding discussion applies. Points of difference are that one diode is reversed, and the output circuit is changed. Resistor R_1 and the condenser to the right of it are large, so that only slow changes of voltage across R_1 can occur.

Assume that the received signal is unmodulated and hence at midfrequency. The two diodes will pass a rectified direct current through R_1 giving it the polarity indicated. This also will charge the large condenser. The combination is so selected that the

Simplified circuit of the ratio detector, a circuit used for demodulating frequency-modulated signals.

voltage across R_1 can vary only in accordance with *slow changes* such as caused by signal fading. Of course if the radio-receiving set is tuned to a strong station, the voltage across the condenser will be large, and if it is tuned to a weak station, the voltage will be small. This makes available an automatic volume-control voltage. If the receiver is tuned to a station radiating a frequency-modulated signal of constant amplitude and if no fading occurs, the voltage across this resistor–condenser combination *will not* change owing to input variations of short duration caused by static or other electric disturbances. This eliminates the need for a limiter (Chap. XIV).

The vector diagrams for the discriminator (Fig. 243) also apply to the ratio detector. If the frequency-modulated signal voltage from plate to cathode of tube 2 exceeds that of tube 1, then condenser C_2 will have a greater rectified voltage across it than will condenser C_1. But the sum of these two voltages must equal the "constant" voltage across R_1. If the frequency-modulated voltage on tube 1 exceeds that on tube 2, then the voltage of condenser C_1 will exceed

that of C_2, but again, their sum must equal the constant reference voltage across R_1. By this action the *ratio* of the voltages across C_1 and C_2 varies in accordance with the changes in the frequency-modulated signal, and an audio voltage is available as indicated.

Phase Modulation.—As has been discussed, phase modulation is one form of angle, or angular, modulation and is closely related to frequency modulation. In a frequency-modulated wave the *instantaneous frequency* is caused to vary in accordance with the magnitude of the modulating audio signal. In phase modulation, the *instantaneous phase* of the output wave is caused to vary in accordance with the magnitude of the modulating frequency. The rate at which the variations occur depends in either type of modulation on the frequency of the modulating signal. A "picture" of a phase-modulated wave would be the same as that of the frequency-modulated wave in Fig. 237. The fundamental difference between phase and frequency modulation is explained[1] as follows: If an audio signal of constant amplitude and variable frequency is used to *phase-modulate* a carrier, the amplitude of the sidebands *stays constant for any value of frequency.* If the same audio signal is used to *frequency-modulate* a carrier, the amplitudes of the sidebands are different for different audio frequencies. Frequency modulation corresponds to a phase modulation where the amplitude of the modulating audio signal is inversely proportional to the frequency of that signal. An application of this principle was given on page 444.

One important difference between phase and frequency modulation is this: For a given audio-modulating signal voltage the radio-frequency band required to handle a phase-modulated signal *is proportional* to the frequency of the modulating signal, but with frequency modulation the radio-frequency band *is independent* of the frequency of the modulating signal. Because of this, a wider radio-frequency band is necessary to accommodate a phase-modulated signal than is required to accommodate a comparable frequency-modulated signal. This statement does not hold for small degrees of modulation.

Methods of Phase Modulation.—These methods are similar to those used for frequency modulation. The system shown in Fig. 239 produces phase modulation if the audio-frequency distorting

[1] See first reference of footnote, p. 443.

network is removed. Other similar circuits are used.[1] In fact, any system of frequency modulation will produce phase modulation if the relative magnitudes of the voltages of each component of the audio modulating signal are made proportional to the frequencies of those components. This can be accomplished by audio-frequency distorting networks. Likewise, systems of phase modulation can be, and are, used to produce frequency-modulated waves.[2]

Demodulation of Phase-modulated Waves.—The circuits used for the demodulation, or detection, of phase-modulated waves may be the same as those used for frequency-modulated waves. However, if a frequency-modulation detector is used, a suitable network must be inserted between the detector and the audio-frequency amplifier to make the amplitudes of the audio components inversely proportional to their frequency.

Phase modulation has been tried experimentally and is used in practice to some extent. At present, it does not have the increasingly wide application of frequency modulation. Although future trends do not seem entirely established, it does not at present appear that phase modulation is to have as wide use as frequency modulation.

Pulse Modulation.—These new systems of modulation already have practical applications. Pulse modulation is obtained when the amplitude or the angle of a carrier wave is "keyed" at intervals. Various forms of pulse modulation are possible[3] among which are the following:

Pulse-time Modulation.—The timing of the pulse relative to a reference pulse is varied around a fixed mean value and conforms to the amplitude of the signal to be transmitted.

Pulse-width Modulation.—The duration of the pulse is varied around a fixed mean value and conforms to the amplitude of the signal to be transmitted.

Pulse-frequency Modulation.—The repetition rate of the pulse is varied around a fixed mean value and conforms to the amplitude of the signal to be transmitted.

[1] See Terman, F. E., "Radio Engineers' Handbook," McGraw-Hill Book Company, Inc.

[2] Mark, M., Cascade Phase Shift Modulator, *Electronics*, Vol. 19, No. 12, December, 1946

[3] "Reference Data for Radio Engineers," Federal Telephone and Radio Corporation.

It appears that the advent of pulse modulation opens a number of new and useful possibilities, and that these pulse-modulation systems are destined to have important uses.

SUMMARY

In a radio system, the audio-frequency signals must be moved or translated to the radio-frequency spectrum. This is accomplished by modulation. After being received, the radio signals must be demodulated or moved back to their original frequencies so that they will be audible.

The types of modulation used at present are amplitude, angle, and pulse modulation. Angle, or angular, modulation is of two types, frequency and phase modulation. Of these, frequency modulation is the more widely used.

In amplitude modulation the modulated wave appears to be a carrier wave that rises and falls in accordance with the modulating frequency. This is because the amplitude-modulated wave is composed of a carrier component and of two sidebands. The information to be transmitted is contained in the sidebands.

To produce amplitude modulation, the carrier to be modulated and the modulating speech or program signal are impressed simultaneously on a circuit that causes nonlinear distortion and creates the desired sidebands. Vacuum-tube circuits commonly are used for this purpose.

In one circuit in extensive use for amplitude modulation the carrier is injected into the grid circuit and the audio into the plate circuit. This is sometimes referred to as plate modulation of a class C amplifier. All the sideband power comes from the audio amplifier. In one type of amplitude modulation both signals are injected into the grid circuit. In the balanced modulator either the carrier or the modulating signal can be eliminated from the output, depending on the connections used. Miscellaneous methods of amplitude modulation include cathode modulation and suppressor-grid modulation.

The beat-frequency oscillator employs the principle of amplitude modulation to create an audio-frequency signal from the output of two oscillators generating much higher frequencies.

In the process of demodulation, or detection, the information in the sidebands is moved or translated back down to frequencies that are audible. This process fundamentally is the same as modulation; it is one of nonlinear distortion in which new frequencies are created—in this instance the audio component.

Crystals are used as detectors. The most widely used demodulator, or detector, is the diode detector. This is good for signals that have a large percentage modulation. Another advantage is that an automatic volume-control voltage is obtainable. This voltage varies in accordance with signal fading, and is impressed on a variable-mu tube in such a way that the output of the radio set is held constant. Triodes also are used as detectors, but to a limited extent in modern radio-receiving sets.

In angle, or angular, modulation, the angle of the impressed carrier signal is caused to vary in accordance with the modulating audio signal. There are

two closely related systems: one is frequency modulation, and the other is phase modulation.

In frequency modulation, the instantaneous frequency of the output wave is controlled by the modulating audio frequency. The magnitude of the frequency excursion, or deviation, from the midfrequency is proportional to the magnitude of the modulating signal. The number of frequency excursions per second depends on the frequency of the modulating signal.

A frequency-modulated wave is a very complex wave. Instead of only two side frequencies, or sidebands, as in amplitude modulation, a frequency-modulated signal contains, theoretically, an infinite number of sidebands. The number of the side frequencies and the magnitudes of these side frequencies are determined by the frequency modulation index, given by Eq. (123)

$$m_f = \frac{f_d}{f_m},$$

where m_f is the modulation index, f_d is the maximum frequency deviation, and f_m is the frequency of the modulating wave.

There are several methods of frequency modulation in common use. One of these is the indirect method used by Armstrong. In this the desired frequency-modulated wave is, in a sense, synthesized. The reactance-tube method, employing ordinary vacuum tubes, also is used. In this scheme, the reactance tube and its circuit act like a reactance across the tuned circuit of an oscillator. The audio signal, when impressed on one of the control grids of the reactance tube, causes the reactance to vary. In this way the oscillator output frequency is caused to vary in accordance with the modulating audio signal. Another system of frequency modulation uses a special tube called a "Phasitron."

Frequency-modulated waves are demodulated in several ways. A typical circuit uses a double diode, and gives out an audio-frequency component when a frequency-modulated wave is impressed. One such circuit is called a discriminator and another is called a ratio detector.

Phase modulation is another form of angle, or angular, modulation in which the phase angle of the signal is caused to vary in accordance with the modulating audio signal. The circuits used for modulation and demodulation are similar to those used in frequency modulation.

Pulse modulation is the latest system used in practice, and shows promise of wide application.

QUESTIONS

1. Why are the low-frequency audio signals changed to radio-frequency signals for transmission in a radio system?

2. What is meant by modulation? By demodulation, or detection?

3. What reasons are there for considering modulation and demodulation to be the same fundamental process?

4. Describe an amplitude-modulated wave in which the modulation is 50 per cent.

5. Can information be transmitted by a radio system if the carrier component is suppressed? Why?

6. Enumerate the types of modulation, and briefly discuss each.

7. Why is it incorrect to assume that the carrier "carries" the information?

8. What reason is there for saying that devices that cause nonlinear distortion can produce amplitude modulation?

9. As used in this book, what is the difference between the modulator and the modulating amplifier?

10. Briefly explain how amplitude modulation by grid and plate injection operates.

11. Why is a tuned circuit placed in the plate circuit of a plate-modulated class C amplifier? What is another and more fundamental name for this device?

12. Briefly explain how amplitude modulation by grid injection operates.

13. Discuss the power-output requirements of the audio modulating amplifier for the modulators discussed in Questions 10 and 12.

14. Are balanced modulators used? If so, what are the important features?

15. Explain how suppressor-grid modulators operate.

16. Briefly explain the theory of the beat-frequency oscillator. Is this term descriptive of the fundamental principle?

17. Give a numerical example of demodulation. Would demodulation result if the carrier and only one sideband were used?

18. Explain the principle of operation of the crystal detector. Is it used in modern radio?

19. Briefly explain how the diode detector produces an audio voltage.

20. Explain how an automatic-volume control voltage is obtained and used.

21. Would the demodulator of Fig. 235 require radio-frequency driving power? Why?

22. What are the characteristics of a frequency-modulated wave?

23. Would a frequency-modulated station require a band in the radio-frequency spectrum that is wider than twice the maximum deviation?

24. Briefly explain how a frequency-modulated wave can be synthesized.

25. What is meant by the frequency-modulation index, and how is it calculated?

26. Briefly explain the reactance-tube method of frequency modulation.

27. Briefly explain what is meant by the disk of electrons in the Phasitron.

28. What is a discriminator, and what is the nature of its output?

29. Very briefly explain important points of similarity and difference for frequency-modulated waves and phase-modulated waves.

30. Is there a similarity between the circuits used in frequency modulation and phase modulation? Discuss briefly.

PROBLEMS

1. The carrier frequency assigned to an amplitude-modulated radio-broadcast station is 620 kilocycles. If a high-quality musical program is being transmitted, what will be the frequency bands occupied by the sidebands?

2. When a certain broadcast transmitter is unmodulated, the carrier output is 5.0 kilowatts. What will be the power in the carrier component when the modulation is 50 per cent and when it is 100 per cent? What will be the power in each sideband for these modulation percentages?

3. A plate-modulated class C amplifier is delivering a carrier of 5.0 kilowatts. About how much direct-current power will the plate circuit draw? How much

audio-frequency power must the modulating amplifier supply at 100 per cent modulation? What type of audio amplifier probably would be used?

4. Calculate the approximate gain of the triode voltage amplifier in Fig. 233.

5. The 150-micromicrofarad condenser and the 220-micromicrofarad condenser of Fig. 233 are to by-pass high-frequency components. Calculate the reactance of each of these to a 500,000-cycle component. Do your results indicate that they would be effective?

CHAPTER XII

RADIO TRANSMITTERS

Radio transmitters include in one form or another many of the circuits and devices that have been discussed in the preceding chapters. A radio transmitter usually is powered by connection to a 60-cycle power source, to a generator, or batteries. In radio telephony the speech, or program, to be transmitted is impressed on the transmitter, and modulates the radio-frequency carrier wave produced by the transmitter. The radio transmitter increases the power level to the desired amount, and impresses the modulated signal on the transmitting antenna, or on the line or cable leading to the antenna. In radio telegraphy, the transmitter is keyed by hand or by automatic sending equipment, and the code-modulated wave is delivered to the antenna.

There are, of course, many types of radio transmitters. Also, the various models made by the different manufacturers have distinguishing features. In this chapter it will be possible to consider only those types which are of fundamental importance. The various component circuits that are used in radio transmitters have been discussed throughout this book. This chapter will explain how these circuits are combined into a transmitter. It will be necessary to limit the discussions mainly to the basic principles involved. No attempt will be made to include wiring diagrams, or to give constructional details. Such information can be found in handbooks, catalogues, and radio journals.

Types of Radio Transmitters.—Those of greatest importance at present are amplitude-modulation transmitters and frequency-modulation transmitters. These two basic types may be subdivided as to their particular field of application.

Radio-broadcast Transmitters.—Amplitude-modulation broadcast transmitters have been used since about 1920. These operate at the familiar broadcast frequencies of about 550 to 1700 kilocycles. Frequency-modulation broadcast transmitters came into use about 1940, and are being installed quite rapidly. The present assignment is at about 100,000 kilocycles, or 100 megacycles.

Point-to-point Radio Transmitters.—Both radio-telephone and radio-telegraph transmitters are used for point-to-point service as contrasted with broadcast. Transoceanic radio-telephone transmitters (and receivers) are used to tie together land telephone lines. Transoceanic radio-telegraph transmitters are used for telegraphic service throughout the world. Point-to-point radio transmitters are used to a limited extent at present for service where land lines might be used, but where it is impossible, or at least extremely costly, to use land lines. Amplitude-modulation, frequency-modulation, and pulse modulation transmitters are used for this purpose.

Mobile Radio Transmitters.—One of the first uses of radio was by ships at sea. In a few instances ship telephone systems are connected by radio to land telephone lines. Radio-telephone service now is given trains, trucks, automobiles, aircraft, and other moving objects. Amplitude modulation and frequency modulation chiefly are used. Phase modulation also has been adopted to some extent.

This gives a few of the various uses of radio and of the types of radio transmitters used. No mention has been made of television and facsimile transmitters, or of many others that might be listed.

Speech-input Equipment.—Before discussing radio transmitters further, the audio-frequency equipment preceding the radio transmitter will be considered briefly. This audio-frequency equipment may be considered as a part of the studio and control room rather than as a part of the radio transmitter.

The types of microphones used in radio, and the acoustical characteristics of the studios in which they are placed, were considered in Chap. I. The electric signals from the microphones are fed into a **control room** in a typical installation. Here are located amplifiers, or **preamplifiers** as they often are called. The cables used for connecting the microphones to the control room are shielded carefully. The types of cables used, and the input impedance to the audio preamplifiers, will depend on the output impedances of the microphones employed. Often these impedances are 50 ohms and 250 ohms. The output level of typical microphones was discussed on page 27. The preamplifiers usually have impedance-matching input and output transformers. These transformers, particuarly the input transformer, must be the type that is carefully shielded, because the signal level from the microphone is very weak. The gain of a typical amplifier is about 50

decibels. Resistance coupling often is used if the preamplifier contains more than one stage.

One microphone is fed into each preamplifier; these preamplifiers in turn feed into a **mixer.** The mixer has this purpose: Sometimes more than one microphone is used in a studio to pick up the program at more than one point. In such instances it is then necessary to mix or bring together the signals from the several microphones. The mixers contain attenuators also, by which the volume from each microphone can be regulated for proper blending together of the signals. The attenuators are placed after the pre-

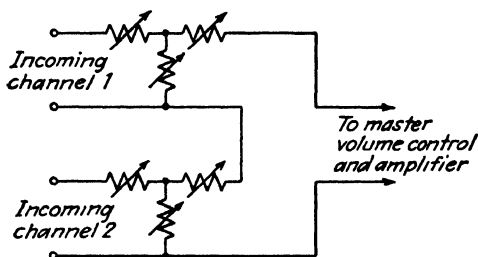
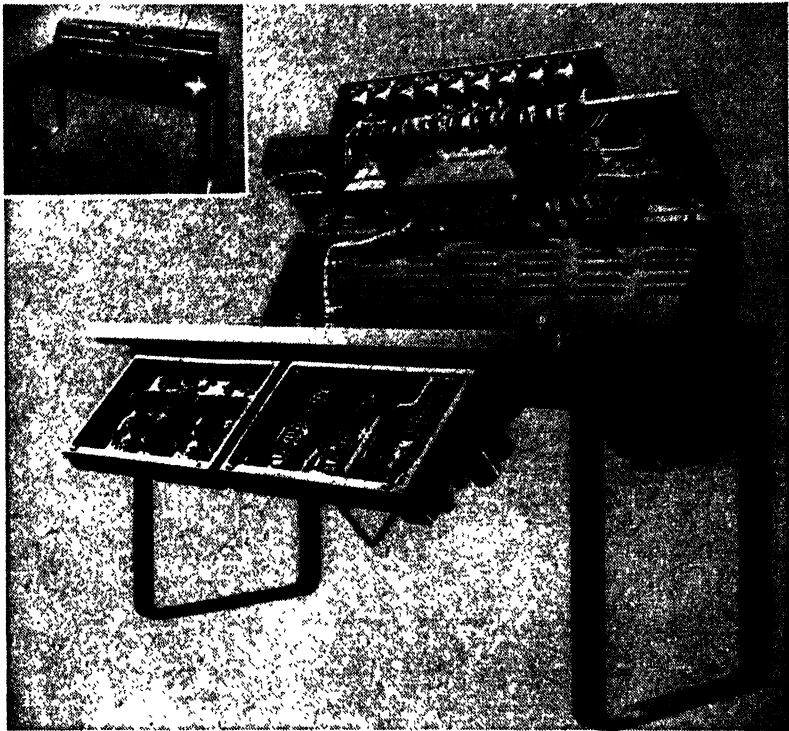



FIG. 244.—One possible mixer arrangement by which the incoming signals from two microphones or other devices can be blended together. The variable resistors are "ganged" together so that varying one dial simultaneously varies all three resistors in each mixer.

amplifier so that the signal adjustments can be made at higher volume level for reasons explained on page 286.

One of the many possible mixer arrangements is shown in Fig. 244. With this circuit either microphone may be used, or the output of the two microphones may be fed at the same time to the transmitter. The mixer of Fig. 244 employs two unbalanced attenuators. These operate in accordance with the theory on page 163. If the microphone lines are balanced, then balanced attenuators must be used. The three variable resistors on a given attenuator are varied by the same knob. If the knob is turned so that the two series resistors are increased, the shunt resistor is decreased, and vice versa. The values are such that the iterative impedance (page 163) is independent of the setting to a large degree. If this is true, then turning one attenuator has negligible effect on the output of the other.

The mixer usually is followed by an amplifier and a master volume control. The output then goes from the control room to the transmitter, which may be at a considerable distance. For

instance, studios and control room often are located in the heart of a city, but the transmitter and antenna usually are located at the outskirts. Various pads, additional amplifiers, and equipment for monitoring the programs often are inserted. Sometimes point-to-point radio is used to feed the program to the remotely located

Speech-input equipment for an amplitude-modulation, or frequency-modulation, radio-broadcast transmitter is shown in this illustration. In the upper left the apparatus is in operating position. The other view is of the equipment opened for servicing. This console-type mounting contains microphone, phonograph, and incoming transmission-line facilities, and provides preamplifiers, attenuators, volume-level indicators, and other control equipment. (*Western Electric Co.*)

transmitter. **Electronic mixing,** such as that shown in Fig. 245, is used sometimes. In this circuit the two portions of the tube operate as a resistance-coupled amplifier.

Amplitude-modulation Radio-telephone Transmitters.—The radio transmitters used in amplitude-modulation systems can be classified on the basis of the point in the transmitter circuit at

which the modulation occurs. In some transmitters the modulation is produced early in the transmitter, where the signals are at a low volume level. This is called **low-level modulation.** In other transmitters the modulation occurs after the signals have been amplified until they are large; this is called **high-level modulation.** These two types of modulation will be explained now.

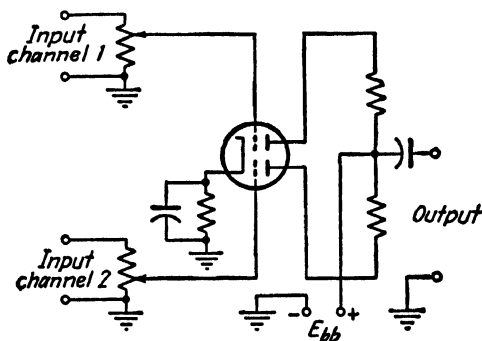



FIG. 245.—An electronic mixer using a double triode.

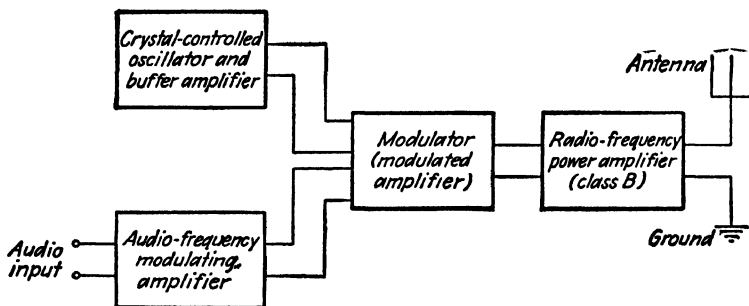



FIG. 246.—Block diagram of an amplitude-modulation radio transmitter employing low-level modulation.

Low-level Modulation in Amplitude-modulation Transmitters.—The arrangement of equipment in a radio transmitter employing low-level modulation is shown in Fig. 246. The circuit details used in a particular transmitter vary considerably, and the following description should be regarded as typical rather than as covering all types.

The crystal-controlled oscillator (page 381) generates the carrier frequency assigned to the transmitter. It is connected to one or more buffer amplifier stages used to increase the volume and

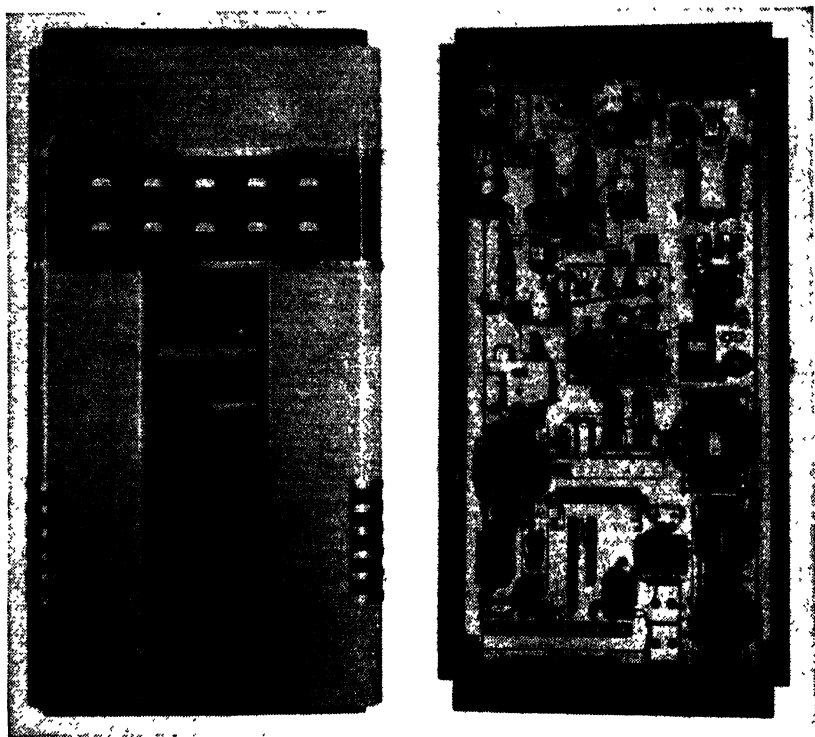
isolate the crystal from the load. Since a single frequency only is to be amplified, the amplifiers associated with the crystal oscillator usually are operated in class C. These amplifiers build up the output of the crystal oscillator until sufficient voltage and power are available for modulation. With low-level modulation, this is but a small per cent of the total power output of the transmitter. In high-frequency transmitters, frequency multipliers may follow the crystal oscillator to give the desired high-frequency carrier wave.

The audio-frequency modulating amplifier receives the audio speech, or program, signal from the microphone and associated speech-input equipment, and increases the signal until it is sufficient to produce modulation. The modulator, or modulated amplifier, often is a tube operated in class C, with both the carrier and the modulating audio signal injected into the grid circuit. When operated in this way, the power required from the crystal oscillator amplifier, and from the audio modulating amplifier, is not great (page 418). A push-pull class A amplifier may therefore be used.

The radio-frequency power amplifier that follows the modulator of Fig. 246 must amplify the carrier and the two sidebands in a broadcast system. Its purpose is to increase the power level of the relatively weak output of the low-level modulator; also, the radio-frequency power amplifier must deliver this power to the antenna. Because the carrier and the sidebands must be amplified, this stage must be operated in either class A or class B; operation in class C would cause excessive distortion when the signal was received and demodulated. Because much power must be handled, and because of the higher efficiency, the radio-frequency power amplifier often is operated in class B. Either one tube or two tubes in push-pull may be used (page 361). This amplifier often is called a **linear amplifier** in radio, because it does not distort the shape of the envelope of the modulated signal (page 350). As previously mentioned (page 407), if this envelope is distorted, then the reproduced speech or music will be distorted.

When a speech or program signal is being amplified, much of the time the modulating signal is low, yet the amplifier must be designed and operated to pass the peak values. This means that during much of the time the class B power amplifier is operating with a low percentage of modulation and with low efficiency, be-

cause the efficiency is grestest when the amplifier is driven to its maximum capacity.

A high-efficiency amplifier has been developed for use as the final radio-frequency power amplifier. This is called the **high-efficiency linear power amplifier, or Doherty amplifier.**[1] In reality, this amplifier consists of two tubes in a special circuit. The first

Front and rear views of a 250-watt amplitude-modulation radio-broadcast transmitter. In the rear view the doors are open. (*Radio Corporation of America.*)

tube is in class B, and works near maximum capacity, and with high efficiency, when the signal is relatively weak. During this interval the second tube passes no signal. When the signal is strong, however, the second tube passes current on the peaks of modulation. The special circuit combines the output of the two tubes so that the output is essentially undistorted, just as if a single

[1] Doherty, W. H., A New High Efficiency Power Amplifier for Modulated Waves, *Proceeding of the Institute of Radio Engineers*, Vol. 24, No. 9, September, 1936.

class B stage had been used. With this combination the efficiency is reasonably high at all times when a speech, or program, signal is being transmitted.

High-level Modulation in Amplitude-modulation Transmitters.—In this system the carrier and the modulating speech, or program, signals are amplified to a high power level before they are impressed on the modulator, or modulated amplifier. The output of this modulated amplifier consists of the carrier and two sidebands, and is fed directly into the antenna, or into the transmission line or cable leading to the antenna. The arrangement of equipment in high-level modulation is shown in Fig. 247.

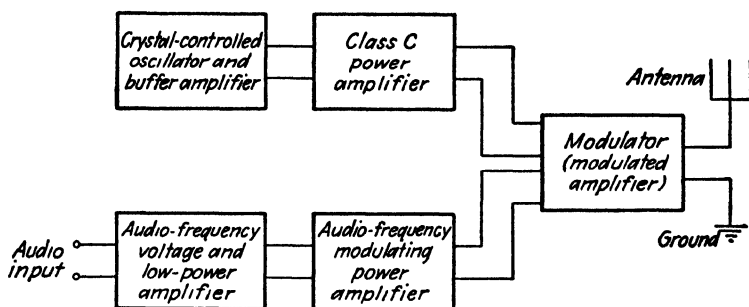



FIG. 247.—Block diagram of an amplitude-modulation radio transmitter employing high-level modulation.

The carrier signal to be modulated is generated in a crystal oscillator provided with buffer amplifiers to isolate it. These feed into a class C power amplifier that increases the carrier to the required power level (page 343). If the crystal frequency is below that required of the carrier, then frequency multipliers may be used.

The speech or program signals to be transmitted are picked up by the microphone, are amplified, and are impressed on the audio-frequency input of the radio transmitter. For high-level modulation, the audio-frequency amplifying equipment may be divided into two portions. The first part is an audio-frequency driver that increases the voltage and provides sufficient power to drive the modulating amplifier (page 412). The final audio-frequency modulating amplifier is a powerful amplifier even in a moderately sized radio transmitter (page 417), and must supply the power that goes into the sidebands for the plate-injection method of modulation (page 414).

A plate-modulated class C amplifier (page 412) usually is employed to produce the modulation in the high-level system. The carrier signal is injected into the grid circuit, and relatively speaking, but a small amount of carrier power is needed. On the other hand, the audio modulating signal is injected into the plate circuit, and must supply the sideband power (page 417).

Miscellaneous Features of Amplitude-modulation Transmitters.—As mentioned at the opening of this chapter, it will not be possible to consider transmitter circuits and operation in detail. There are, however, certain important features that should be discussed, if only briefly.

Neutralization.—The circuit arrangements for accomplishing neutralization were discussed on page 356. When a transmitter is placed in operation, all neutralized stages must be adjusted so that the neutralization is effective. A simple procedure for adjusting the neutralization of a radio-frequency amplifier is as follows: With the tube or tubes (if in push-pull) in normal operation, apply the carrier-frequency alternating voltage, and adjust the tuned input and output stages to parallel resonance. There are various ways of noting when resonance occurs, for example, by observing the alternating-current flow, by observing the alternating voltage, or by bringing a wavemeter up to the circuit being adjusted. The **wavemeter** is a calibrated series resonant circuit with a thermogalvanometer to indicate maximum current flow, and a maximum deflection indicates resonance of the circuit being tuned. The plate supply is then removed from the tube being neutralized, and the tuning again is adjusted for resonance. The neutralizing condenser is then adjusted so that no alternating current flows in the tuned plate circuit and no alternating voltage exists across it. Under these conditions a signal in the grid circuit causes *no* signal in the plate circuit, and for this adjustment a signal in the plate circuit (when the tube is operating) will feed back *no* signal to the grid circuit. This is an application of what is known as the **reciprocity theorem.**

Parasitics.—What are known as parasitics, or more properly **parasitic oscillations,** are likely to occur in large radio-frequency oscillators and amplifiers. These are undesired oscillations, and they may cause reduced power output at the desired operating frequency, or distortion or radiation of undesired signals, or they may even cause flashover of equipment. This subject is very

involved, because parasitic oscillations may occur in many ways.[1] As an illustration of *one of many types* of parasitic oscillation, consider the tuned-grid tuned-plate oscillator of Fig. 248*a*. At very high frequencies, the actual oscillator becomes equivalent to the circuit of Fig. 248*b*, because at very high frequencies the tuning condensers in the grid and plate circuits have negligible reactance. The circuit of Fig. 248*b* is that of a very high frequency oscillator having an inductance of a single turn to resonate with the interelectrode capacitance. In the design of new equipment, parasitic oscillations are to be expected. They must be hunted out with a small neon bulb, or a search coil, or with a wavemeter, or in other

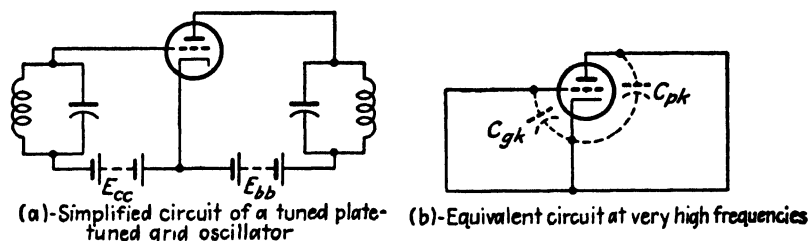



Fig. 248.—At very high radio frequencies the reactance of the tuning condensers becomes negligible and the equivalent circuit of (*b*) results. Parasitic oscillations at very high frequencies may occur in this circuit.

ways. A more or less experimental technique is required to remove the parasitic oscillations. Sometimes leads are shortened, wiring is rearranged, and equipment is changed in position. Sometimes resistors or chokes are added to leads, and sometimes they are removed from leads. There are many causes and many remedies for parasitics.

Overmodulation.—An amplitude-modulation transmitter should be so adjusted and operated that for the peak of the modulating wave the modulated output composed of the carrier and sidebands is just reduced to zero. This is indicated in Fig. 249*a* when the modulating signal is a pure sine wave, and will be recognized as 100 per cent modulation. If the peak value of modulating signal is too great, then the percentage modulation will exceed 100 per cent, overmodulation will result, and the signal will be distorted. This will cause at least two bad effects: (*a*) when the received signal is received and demodulated, it will be distorted; and (*b*) the radiated signal will contain frequency components other than the

[1] An excellent treatment is given by F. E. Terman, "Radio Engineers' Handbook," McGraw-Hill Book Company, Inc.

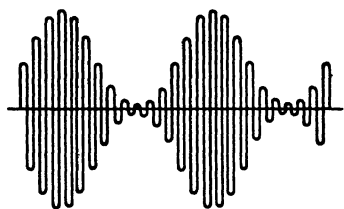
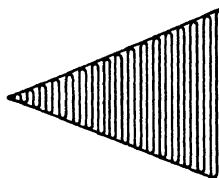
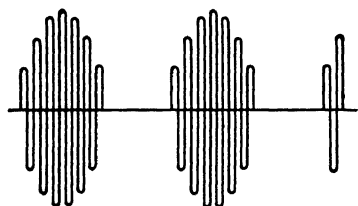
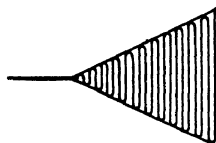
carrier and the two sidebands, and interference with other amplitude-modulation radio channels will result. Special instruments are available for determining when overmodulation occurs. A cathode-ray oscilloscope (page 575) also can be used for this purpose: The modulating signal voltage is applied to the horizontal deflecting plates, and the modulated output voltage, composed of the carrier and sidebands, is applied to the vertical deflecting

(a)-Amplitude-modulated wave with 100 per cent modulation

(c)-Cathode-ray oscilloscope pattern for 100 per cent modulation

(b)-Overmodulated amplitude-modulated wave

(d)-Cathode-ray oscilloscope pattern for an overmodulated wave

FIG. 249.—Approximate characteristics of amplitude-modulated waves when the modulating signal is a pure sine wave such as 1000 cycles.

plates. If the modulation is 100 per cent, the figure on the end of the cathode-ray tube will appear as in Fig. 249*c*. If the modulation exceeds 100 per cent, the signal will appear as in Fig. 249*d*. Furthermore, amplitude distortion in the modulator is indicated if the upper and lower sides are curved instead of straight lines, and phase shift is denoted if the figure is elliptical. The modulating signal in Fig. 249 is a sine wave, but in practice it is speech or music with decided peaks, and overmodulation on strong peaks is difficult to prevent. Thus, the speech-input equipment to radio-broadcast transmitters (and other transmitters as well) often contains an audio-frequency **limiting amplifier** so designed that excessive peaks

are suppressed. With this arrangement, the average percentage modulation can be much higher without the danger of overmodulation. This means, of course, that a stronger over-all signal can be radiated. The signal distortion, called **clipping**, caused by suppressing the peaks is not noticeable.

Negative Feedback.—This is used in amplitude-modulation radio transmitters to reduce distortion and noise; the basic principle is presented in Fig. 250. The theory essentially is the same as when feedback is used with amplifiers (page 361). A distinguishing fea-

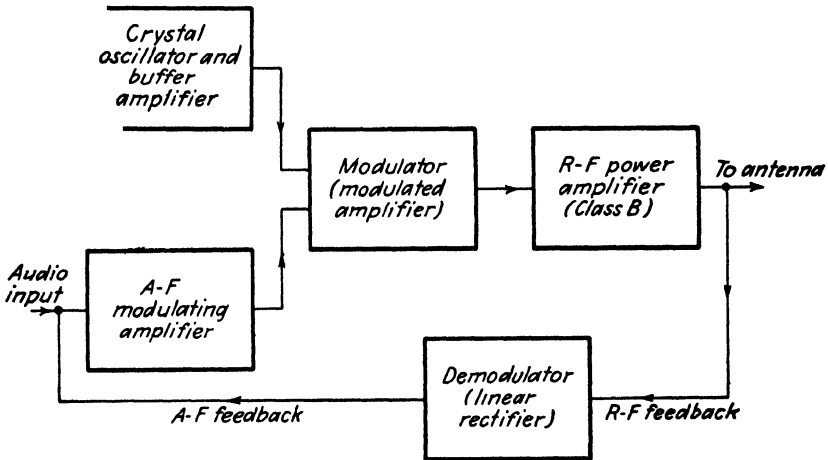
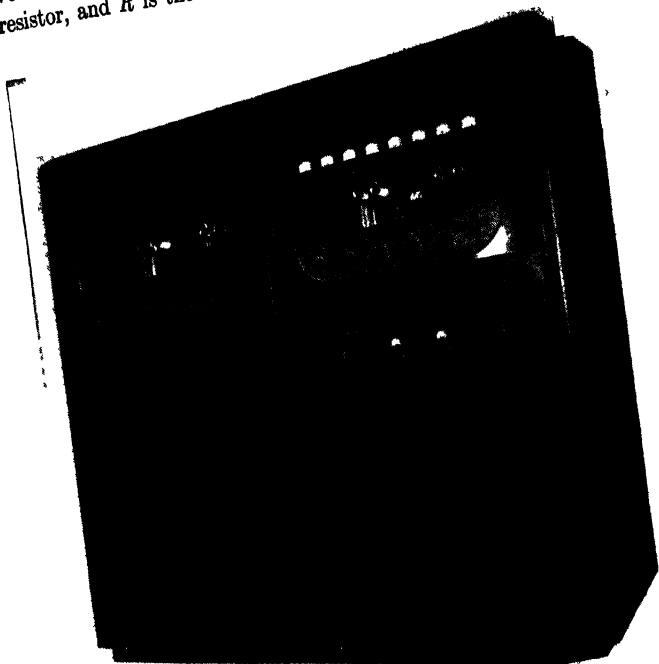



FIG. 250.—Negative feedback applied to a low-level amplitude-modulation transmitter.

ture is that a demodulator (often called a "linear rectifier") must be placed in the feed-back circuit to obtain an audio-frequency component to feed into the system with the incoming audio signal (see page 426). Feedback will not prevent distortion caused by overmodulation.

Dummy Antennas and Power Measurements.—When preliminary adjustments are made on a radio transmitter, it is important that it is not "on the air" and radiating radio-frequency energy. The transmitter must be loaded, however, just as it will be loaded in actual operation by the transmitting antenna. Some transmitters are provided with a **dummy antenna** which will load the transmitter and dissipate energy, but which will not radiate appreciable energy into space. A dummy antenna may be a resistor arrangement or may be incandescent lamps. They must be connected to the trans-

mitter or coupled into the transmitter output so that they simulate actual conditions. The power dissipated, which is the output of the transmitter, can be computed by E^2/R or I^2R, where E is the voltage across the load resistor, I is the current through the load resistor, and R is the radio-frequency resistance at the operating

Front view of an amplitude-modulation radio-broadcast transmitter. This set will deliver either 500 or 1000 watts to the antenna. (*Collins Radio Co.*)

frequency and temperature. The voltage can be measured with a vacuum-tube voltmeter (page 351), and the current with a thermo-ammeter (page 91). If these instruments are not available, and if lamps are used as the dummy antenna, there is a simple approximate method: Arrange a phototube, photocell, or light meter to measure the intensity of the light when the radio transmitter is feeding energy to the lamps; then disconnect the transmitter and supply the lamps from a direct-current source or from a 60-cycle

supply, and adjust to the same light intensity as before. The power as determined from direct-current ammeter and voltmeter readings, or from 60-cycle wattmeter readings, will equal approximately the radio-frequency output of the transmitter.

Transmitter Coupling Networks.—The parallel tuned circuit of the last power-output tube of a transmitter must be coupled to the antenna, or to the line or coaxial cable feeding the antenna, by a network that reflects the correct impedance into the parallel tuned circuit. If this is not done, the power tube will not be properly loaded, and correct power transfer conditions will not be obtained. Before this can be discussed, it is advisable to know the input-impedance characteristics of transmitting antennas, and the subject will be deferred until the next chapter.

Single-sideband Amplitude-modulation Radio Telephony.—At various points in the preceding chapters it has been mentioned that in amplitude modulation *all* the information or program to be transmitted existed in *each* sideband. It also was explained that in a 100 per cent amplitude-modulated wave two-thirds of the output power and output power-handling capacity was required by the carrier, and that only one-third was devoted to the two sidebands. Also, it was stated that single-sideband communication was possible and was used. It is of importance to note that a single-sideband system would occupy only one-half the space in the radio spectrum as compared with the double-sideband system.

It is in special point-to-point systems used for communication (as distinguished from broadcast) that the single-sideband systems are used. Of course, when the carrier is not transmitted, a wave having the frequency of the missing carrier must be injected into the demodulator of the distant receiving set for demodulation or detection, but this can be furnished by an oscillator incorporated in the distant receiving set.

The methods by which the carrier and one sideband are eliminated in these special systems is discussed in several papers[1] (see also page 421). In the early long-wave low-frequency system, the carrier is suppressed entirely. In the later short-wave system, it is transmitted at greatly reduced volume and is used at the distant

[1] Heising, R. A., Production of Single Sideband for Transatlantic Radio Telephony, *Proceedings of the Institute of Radio Engineers*, Vol. 13, No. 6, June, 1925. Polkinghorn, F. A., and N. F. Schlaak, A Single-sideband Short Wave System for Transatlantic Telephony, *Proceedings of the Institute of Radio Engineers*, Vol. 23, No. 7, July, 1935.

receiving station as a control frequency for the oscillator that generates the carrier used in demodulation.

Amplitude-modulation Radio-telegraph Transmitters.—The previous pages have dealt with radio telephony in which the modulating signal to be transmitted is speech, or a musical program. This section will consider radio *telegraphy*, where the signal consists of a code of dots and dashes, such as would be sent by a hand-operated telegraph key, or by an automatic sending device.

There are two systems of radio telegraphy in practical use. In the *first system*, the radio-telegraph transmitter gives out "spurts" of high-frequency energy when the transmitter is keyed, and these impulses correspond to the dots and dashes of the telegraph code. This system is used extensively. In the *second system*, radio equipment that is similar to a radio-telephone transmitter is modulated by a signal, such as 1000 cycles, and this modulating signal is keyed in accordance with the telegraph code.

In either system, sidebands are created, and a band of frequencies, instead of a single frequency, is required for transmission through space. In general, the bands required are not so wide as in radio telephony, and in general, the transmitters used in radio telegraphy are simpler than for radio telephony.

A dot and a dash of the code signals that would be produced by the hand-operated telegraph key are shown in Fig. 251*a*. In Fig. 251*b* are shown the shapes of the high-frequency waves that would be sent out *if* the current were allowed *immediately* to build up to its maximum value, and immediately to die out to zero. If this were done, the radiated signal would contain many harmonics, and these would cause the radiated signal to cover a wide frequency band. There is a reason for this: As was explained on page 88, a wave that suddenly starts and stops contains many harmonics of the fundamental. Thus when code signals with "square corners" are radiated, these would not radiate a wave of a single frequency, but would radiate a fundamental and harmonics that would cover a wide band.

Radio-telegraph Interference.—If the signals radiated by a radio-telegraph transmitter are shaped as in Fig. 251*b*, the harmonics will cause interference with other radio-communication channels. Since the harmonics are caused by the fact that the signals start and stop abruptly, the remedy is to "round off" the signals, as indicated in Fig. 251*c*. Tests have shown that radio-telegraph

signals are easier to receive when the "make" end, or starting of the signal, is rather abrupt and the "break" end, or stopping of the signal, is somewhat gradual. This is shown in Fig. 251*c*. A rather abrupt rise causes a slight "click" to be made upon reception, and this assists in distinguishing between the signals. As previously mentioned, however, the build up, or rise, of the current should not be too abrupt or interference, called **key clicks,** will be caused with other radio channels.

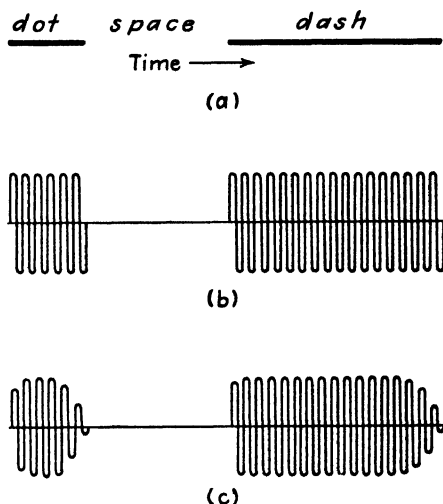



FIG. 251.—Shapes of signals in a radio-telegraph system. The signals are "rounded off" before radiation so that key clicks and interference with other systems are minimized.

Keying Circuits.—In the most widely used type of radio-telegraph transmitter, when the telegraph sending key is open, no signal is radiated, and when the key is closed to send a dot or a dash, "spurts" of radio-frequency energy are transmitted. In a sense, the keying circuit of a radio-telegraph transmitter corresponds to the modulating circuit of a radio-telephone transmitter.

Because a single frequency only is to be amplified, class C amplifiers are used throughout many radio-telegraph transmitters. Quartz-crystal oscillators usually are employed to supply the basic carrier frequency. For very high frequencies, frequency multipliers may be used, or transmission lines instead of crystals may be used to control the frequency of oscillation (page 388).

The systems of keying, or causing the transmitter to send "dots"

and "dashes" of radio-frequency energy, such as the ones shown in Fig. 251, are many and varied. In general, it is most satisfactory to key the transmitter as near to the crystal oscillator as possible; sometimes the oscillator itself is keyed. The reason for this is that if keying is done at a high level (near the output), then all preceding portions are in operation at all times, and this reduces the power efficiency. Furthermore, unless the transmitter is well designed and carefully shielded, undesired radiation will occur when the key is not depressed. This radiation is called a **back wave.**

A method of keying is shown in Fig. 252*a*, in which the keying occurs in the primary of the transformer which supplies alternating-current power to the transmitter rectifier. Actually, the hand-operated key itself would operate a relay, and the relay would in turn operate the power-handling contacts *S*. With this method, the transmitter would be turned on and off in accordance with the code signals. For obvious reasons this method is limited to small transmitters. Another method is shown in Fig. 252*b*. In this the "key" (which is for safety and convenience a relay controlled by the hand-operated key) is placed in the high-voltage direct-current supply to the transmitter. A common method of keying is shown in Fig. 252*c*, and this is known as **cathode keying.** An alternate method is shown in Fig. 252*d*, and it is known as **center-tap cathode keying.** These keying circuits sometimes are placed in the cathode of the oscillator tube, but perform somewhat better when placed in one of the class C amplifier tubes. With cathode keying, when the cathode circuit is opened, the cathode potential rises to a high positive value, because the electrons that leave the cathode and flow to the plate cannot return to the cathode, since the circuit is open. When the cathode rises to a high positive potential, this is equivalent to placing a high negative potential on the grid, and the tube is blocked.

A **lag circuit** is shown in Fig. 252*b*, and similar lag circuits are used often with the keying circuits of Fig. 252*c* and *d*. The purpose of these lag circuits (which are in reality filters) is to slow down the increase and decrease of current and to round off the corners, as shown in Fig. 251. This suppresses the harmonics and key clicks. In this connection it should be recalled that a condenser tends to maintain the voltage across a circuit at a constant value, and a coil tends to maintain the current through a circuit at a constant value.

As has been explained, in many radio-telegraph circuits when the key is depressed a radio-frequency wave containing much

(a)-Keying of the a-c power source

(b)-Keying of the d-c power supply

(c)-Cathode keying

(d)-Center-tap cathode keying

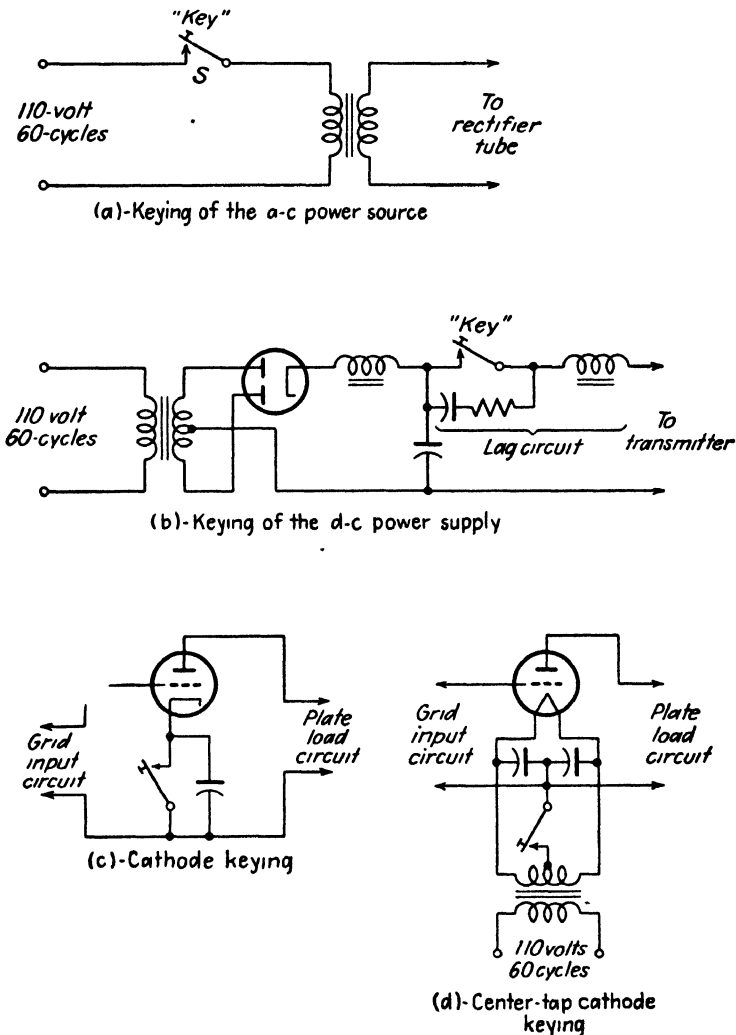



Fig. 252.—Methods of keying radio-telegraph transmitters. In most instances the "key" shown actually would be a relay operated by the sending key. This is for safety and flexibility.

energy is radiated, and when the key is open no signal (or almost none) is radiated. Keying a radio-telegraph transmitter places sudden load variations on the direct-current power supply, which

usually is a tube rectifier. These sudden power-load variations may cause switching transients in the rectifier, and these in turn cause transient output-voltage variations, which may distort the shapes of the transmitted impulses. This may be minimized by careful design.[1]

Frequency-modulation Radio-telephone Transmitters.—Amplitude modulation has been the method used in radio for many years, and because of this, the transmitters used are somewhat standardized. By contrast, frequency modulation is new in application, and the transmitters are not standardized. In fact, quite the opposite is true; and it undoubtedly will be some time (if ever) before the designs are standardized to the extent that has been reached in amplitude modulation.

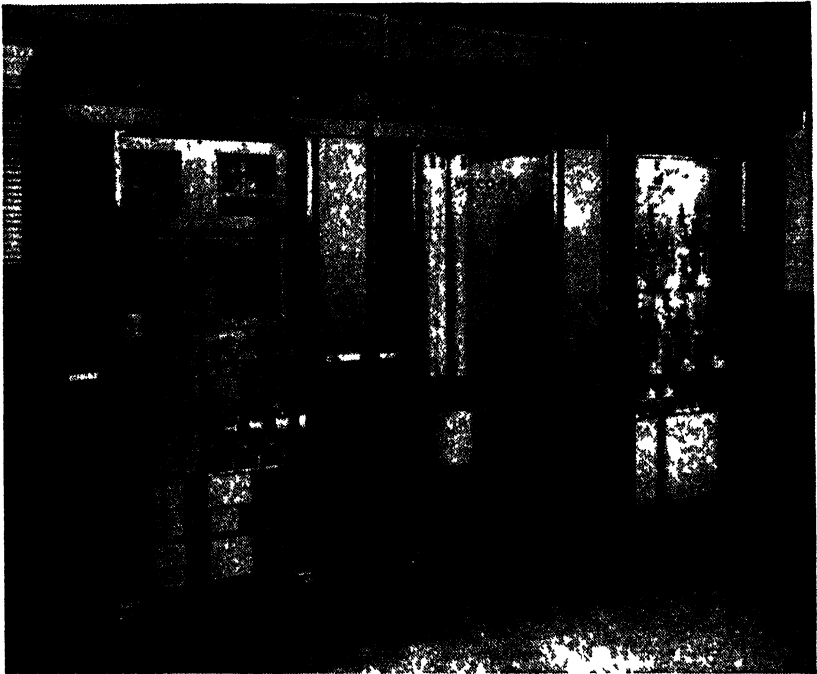
At the present, each of the various manufacturers of frequency-modulation transmitters has a transmitter differing somewhat, and often greatly, from that of the others. In general, the differences are in (*a*) the method of producing the frequency modulation, and (*b*) the method of frequency stabilization employed to hold the midfrequency (or unmodulated frequency) at the assigned value.

The nature of a frequency-modulated wave was discussed in Chap. XI, and several typical methods of producing frequency modulation were also discussed. The methods were (*a*) the scheme by which a frequency-modulated wave was synthesized, or built up, by various steps involving a balanced amplitude modulator and phase shifting; (*b*) the reactance-tube method in which a tube and its associated circuit acted as an audio signal-controlled reactance across the tuned circuit of an oscillator and thereby caused the oscillator frequency to vary in accordance with the speech signal, or program, to be transmitted; (*c*) frequency modulation by means of the Phasitron. Methods of obtaining frequency-modulated waves by phase-modulation processes were also mentioned.

The methods of frequency stabilization have not been discussed. In this respect transmitters for frequency-modulation systems differ considerably. There are at present about a dozen manufacturers of frequency-modulation broadcast equipment, and space limitations prevent the presentation of all systems. Several typical transmitters will be discussed, using block diagrams.[2] In

[1] Terman, F. E., "Radio Engineers' Handbook," McGraw-Hill Book Company, Inc.

[2] The diagrams are based on a chart entitled "Frequency Modulation Systems," issued as a supplement to the journal *Electronic Industries*, April, 1946.

selecting three of the systems for presentation, it is pointed out that the selection is based largely on ease of explanation. Each of the systems must meet the specified stability requirements or they could not be used. The present requirements are that the unmodulated or midfrequency must remain within plus or minus 2000 cycles at a midfrequency of 100 megacycles.

Front view of a frequency-modulation radio-broadcast transmitter, designed to operate in the 100-megacycle region Glass doors permit viewing apparatus during operation This transmitter has an output of 10 kilowatts, and employs "coaxial-cable" tuning of the power amplifier shown in the center The speech-input equipment and the modulating circuits are in the unit at the left The high-voltage power supply equipment is at the right (*Western Electric Co*)

The block diagram of a frequency-modulation transmitter is shown in Fig. 253. It will be recognized as employing the method of frequency modulation by wave synthesis, which was explained on page 443. The basic carrier frequency is controlled directly by a quartz crystal. It will be noted that the phase shifter in Fig. 253 is placed ahead of the balanced modulator, instead of after it as shown in Fig. 239, page 444. With this system, the frequency at which the wave is synthesized is low, and the modulation index

also is low. In early systems, this was of the order of a few hundred thousand cycles and an index of unity. It is necessary to increase the frequency to the bands assigned to frequency-modulation stations. These bands at present are in the vicinity of 100 megacycles. Also, it is necessary to increase the modulation index

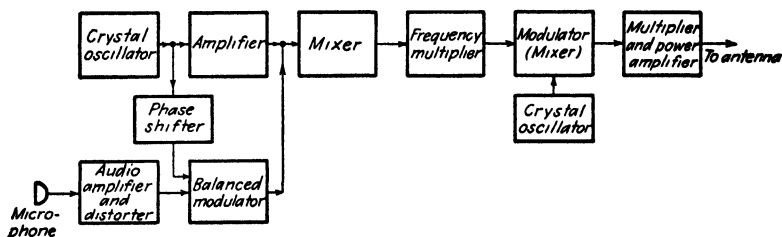



FIG. 253.—Block diagram of a frequency-modulation radio-broadcast transmitter employing the wave-synthesis method. For details of this system, see page 444. The second crystal oscillator and the associated modulator are for raising the frequency-modulated signal to the very high-frequency bands assigned to frequency modulation. This principle is fundamentally the same as explained on page 409. The first mixer merely combines the signals from the two paths into the desired synthesized frequency-modulation wave.

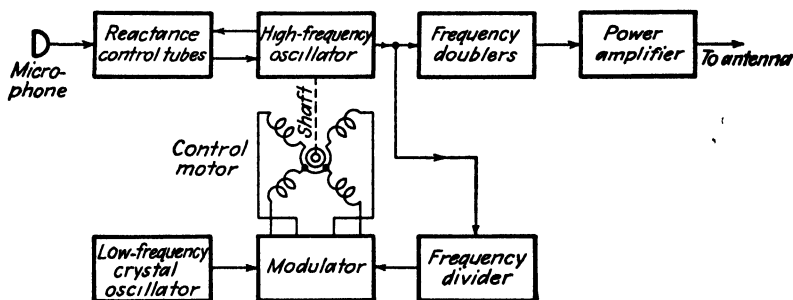



FIG. 254.—Block diagram of a frequency-modulation radio-broadcast transmitter using reactance tubes for modulation and a motor for adjusting the midfrequency. The unit marked "modulator" is part of the frequency-control system, and is not the program modulator.

so that the advantages of noise reduction will be realized (page 568). This is done by the frequency multipliers and by the mixer-oscillator combination in the right-hand portion of Fig. 253.

Several frequency-modulation transmitters use a motor-control arrangement to maintain the midfrequency constant; one system of this type is shown in Fig. 254. The shaft of a special control motor is connected to the condenser of the tuned circuit of the main oscillator, or **master oscillator** as it is often called. The

output wave passes through **frequency dividers,** which reduce the frequency in exact multiples. These consist of a copper oxide bridge arrangement and a vacuum tube.[1] The "modulator" shown in the lower part of Fig. 254 is not the device that produces the frequency modulation; this is produced by reactance-control tubes (page 444) connected to the master oscillator. It is the audio-frequency variations in reactance that cause the frequency of the master oscillator to vary and to produce the frequency-modulated wave. It will be noted that the divided, or subharmonic, frequency and the output of a very low-frequency crystal oscillator are fed into the so-called "modulator," a term

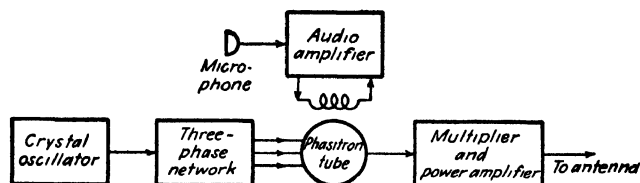



FIG. 255 —Block diagram of a frequency-modulation broadcast transmitter using the Phasitron tube (page 446).

used in a general sense. The output of this goes to the special motor mentioned before. If the midfrequency of the master oscillator tends to rise, the motor is caused to rotate in one direction and to alter the capacitance so that the master oscillator assumes the correct frequency. If the frequency tends to fall, the motor revolves in the opposite direction and corrects the frequency The crystal oscillator produces the reference frequency against which the divided midfrequency is compared and corrected.

The frequency-modulation transmitter of Fig. 255 employs the Phasitron tube. The basic frequency is generated in a crystal oscillator, and is changed into a three-phase voltage by special networks; and this voltage is used as explained on page 447. The carrier frequency is of the order of a few hundred thousand cycles, a frequency at which crystals are very accurate. The frequency and modulation index of the output of the Phasitron is too low, and the output wave passes through several stages of frequency doublers and triplers before reaching the antenna. With this method,

[1] For a discussion of these, see J. F. Morrison, A New Broadcast-transmitter Circuit Design for Frequency Modulation, *Proceedings of the Institute of Radio Engineers*, Vol. 28, No. 10, October, 1940.

the output midfrequency is an exact multiple of the frequency of the crystal.

Television.—Although this book is devoted to the fundamentals of radio, it seems appropriate at this point to discuss television, even if space limitations require that the treatment be brief. By a system of television the image of an object, or a scene, is transmitted, in effect, from one point to another. The speech or sounds accompanying the object, or scene, also are transmitted and are reproduced at the distant location at which the transmitted image is viewed.

A system of television requires several distinct steps or processes as follows:

a. The object or scene must be "scanned" electrically by a **television camera.** This device continuously scans or "studies" in sequence each small section, or element, of the image to be transmitted. As the scanning occurs, the camera gives out an electric current, the strength of which is proportional to the light reflected from each element of the image scanned.

b. These electric signals are amplified and used to amplitude-modulate a high-frequency carrier, and sidebands are created. As will be evident later, the band width is very great, and hence one sideband is suppressed usually.

c. The television impulses, and certain control impulses, are radiated through space. (In early experimental systems, television was accomplished over wire lines.)

d. At the distant station, the television signals are received, demodulated, and are used to operate a cathode-ray tube. The control impulses mentioned in (*c*) are to ensure that the impulses actuating the cathode-ray tube are in exact synchronism with the impulses produced by the camera. In a sense, the cathode-ray beam "paints" the image, or scene, on the end of the cathode-ray tube.

A consideration of a system of television includes devices, such as the television camera and the cathode-ray tube, that have not been discussed as yet. These will be considered in the following pages, which will include also a step-by-step explanation of a television system. In studying this material, it should be kept in mind that television is just emerging from the early experimental stage, and that in all probability it will undergo extensive changes.

The Television Camera.—The device that "studies" the object,

or scene, to be televised is of much importance. In early systems a rotating scanning disk and a phototube were used. Later systems employed the **image dissector** and the **Iconoscope.** A recent television camera uses the so-called **Image Orthicon** tube, a special electronic device of great promise. A television camera employing this tube will be discussed.

The Image Orthicon tube involves many important principles, and a complete discussion is quite detailed.[1] The practical operation of this tube has been summarized in an interesting article.[2]

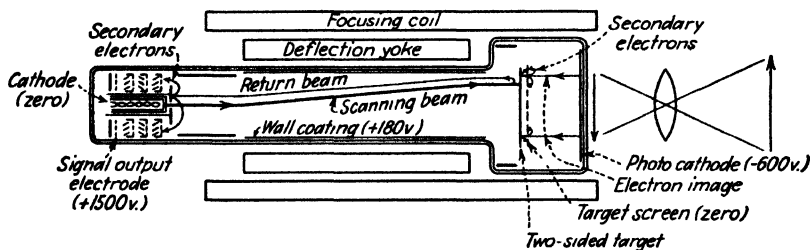



FIG. 256.—Diagram of the Image Orthicon, a television camera tube of much promise. (*American Institute of Electrical Engineers.*)

A schematic diagram of the Image Orthicon is shown in Fig. 256. The photo cathode is coated with a material that releases electrons when light strikes it. The light reflected by the object, or scene, to be televised is focused on the photocathode. Areas that are strongly illuminated give off many electrons, but areas that are weakly illuminated give off few electrons. In this way an "electron image" is formed. The photo cathode is at −600 volts with respect to the two-sided target and target screen. Thus an electron current corresponding to the electron image flows to the target and screen. The image to be transmitted exists in electrons in any cross section of this electron current. Or in other words, at certain small areas in the cross section many electrons, or few electrons, will exist, depending on the "lights" or "shadows" of the optical image. The electrons traveling from the photocathode to the target and screen are held in parallel paths by the magnetic field of the focusing coil.

[1] Such a discussion is given in an article by A. Rose, P. K. Weimer, and H. B. Law, The Image Orthicon—A Sensitive Television Pickup Tube, *Proceedings of the Institute of Radio Engineers*, Vol. 34, No. 7, July, 1946.

[2] Kell, R. D., and G. C. Sziklai, Image Orthicon Camera, *RCA Review*, Vol. VII, No. 1, March, 1946.

The so-called "two-sided target" is of low-resistance semiconducting glass. It is so thin that charges flow readily from one side to the other, but do not flow sideways or laterally during a complete scanning period (to be described later). The electrons flowing from the photo cathode strike one side of the glass target, and because of their high velocities they knock secondary electrons out of the glass target. More secondary electrons are emitted than there are striking primary electrons, and a positive charge pattern is produced on one side of the glass target. The various areas of this positive charge pattern will correspond to the lights and shades of the object, or scene, being televised.

A beam of electrons from an electron gun at the opposite end of the Image Orthicon tube is directed at the target. This beam scans the side of the glass target opposite to that on which the electron image impinges. This beam of electrons from the electron gun is caused to move in lines across the tube, much as a person scans the page of a book, reading a line across the page, hurriedly shifting the eyes back, and then reading the next lower line.

Because the glass is somewhat conducting, if an area on one side is positive, because primary electrons knock out secondary electrons, the area on the reverse side is positive. If there is no positive charge on the target, all the electrons in the scanning beam are reflected back as a return beam. If there is a positive charge on a given area, some electrons are taken from the scanning beam, and fewer electrons are returned back. In this way an electron beam is in a sense modulated or controlled in accordance with the lights and shadows of the object, or scene, to be televised.

The electron gun, previously mentioned, consists of a thermionic cathode that emits electrons and of an arrangement for accelerating and focusing the electrons into a small beam, or ray. It is this ray that goes out and is reflected as a modulated return beam. The return beam strikes a comparatively large electrode around the aperture from which the beam originally was shot. The return beam produces secondary electrons when it strikes this electrode, and this electron current then enters additional electron-multiplying stages, which are in principle like the electron multiplier described on page 212. From the output stage of the electron multiplier the television signal enters a video or wide-band amplifier (page 299), and is ready to be impressed on the television transmitter. Note that at each instant the strength of the current is

determined by the light reflected from a given small area of the object, or scene, to be televised.

Television Transmitters.—As previously mentioned (page 299), for a good "picture," the television signal must occupy a wide frequency band, from about 30 to 3,000,000 or 4,000,000 cycles. This is the **video signal** that corresponds to the audio signal in a radio-telephone system.

This video signal is amplified and used to amplitude modulate a high-frequency carrier. Several television channels are available, certain ones being in the vicinity of 100 megacycles. Special high-frequency tubes have been developed for television signal trans-

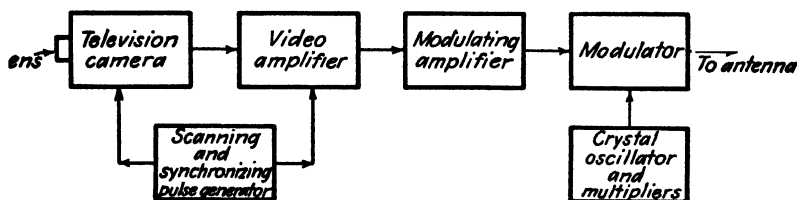



FIG. 257.—Block diagram of a television transmitter.

mitting purposes. Because of the wide frequency band required by the video signal, the band occupied after modulation in the radio-frequency spectrum is wide also. Thus in some systems one sideband is suppressed. A block diagram of the "picture" portion of a television transmitter is shown in Fig. 257.

The speech or program sounds accompanying the scene to be televised are picked up by a microphone in the usual manner, amplified, frequency-modulated, and transmitted by frequency-modulation radio to the distant receiving station.

The way in which the received radio-frequency television signals are demodulated, or detected, and used to reproduce, or "paint," the image, or scene, at the distant station will be discussed in Chap. XIV.

SUMMARY

The speech, or program, signals to be transmitted are impressed on the speech-input circuit of a radio transmitter, and modulate the carrier. This translates the signals to the radio-frequency range. The radio transmitter increases the power level of the signals and impresses them on the transmitting antenna.

The types of radio-transmitting sets of greatest importance are (*a*) amplitude-modulation transmitters and (*b*) frequency-modulation transmitters. Both these types are used for broadcast purposes. For point-to-point service,

radio transmitters employing amplitude modulation, frequency modulation, and pulse modulation are used. Transmitters of these types are used also for mobile installations, with amplitude modulation and frequency modulation being the most common.

In radio telephony, the microphone picks up the sounds to be transmitted and gives out corresponding electric signals. These are amplified in the speech-input equipment, and are mixed and so controlled that the transmitting equipment is not overloaded.

Low-level modulation, followed by radio-frequency power amplifiers, is used in some radio-telephone transmitters. In high-level modulation, the carrier and the modulating audio-frequency signals are amplified to a high level before modulation occurs.

If a radio transmitter uses triodes, the neutralization must be adjusted before the transmitter is put in service. Parasitic oscillations are likely to occur in a transmitter of new design. These oscillations must be eliminated to a great extent by a cut-and-try process. When a transmitter is in operation, the modulating audio-frequency signal must not be so great that overmodulation occurs. Limiting amplifiers are used sometimes to prevent overmodulation. Negative feedback is used with modern broadcast transmitters to reduce noise and distortion.

Single-sideband transmitters are used for special transoceanic radio-telephone services, but are not used for radio broadcast. In the single-sideband system, the carrier frequency that is required for demodulation is generated at the receiving station. In some systems, the carrier is transmitted in reduced magnitude and is used to control the oscillator at the distant receiving station.

In the system of radio telegraphy most widely used, when the sending key is depressed, "spurts" of high-frequency power of the carrier frequency are transmitted. These correspond to the dots and dashes of the telegraphic code. In a common system, the keying circuit is placed in the cathode circuit of the oscillator tube. Lag circuits, which in reality are filters, are used to "round off" the dots and dashes as they start and stop, so that harmonics causing key clicks are reduced. The sudden application of the load when a transmitter is keyed may cause switching transients in the power supply.

Partly because of its newness, frequency-modulation has not become standardized, and transmitters operating on quite different principles are in use. In general the differences are in (*a*) the method of producing the frequency modulation, and (*b*) the method of frequency stabilization employed. The reactance-tube method is the most widely used in frequency modulation; other methods include the wave-synthesis method and the Phasitron. A quartz crystal is used in all frequency-modulation broadcast transmitters to stabilize the frequency, although the details differ widely.

A system of television uses a television camera to scan the image to be tranmitted. Electric waves corresponding to the "lights" and "shadows" of the image are given out by the camera. These electric signals are amplified and are then used to amplitude-modulate the television transmitter. At the distant station, the received radio signals are demodulated and used to control a cathode-ray tube the beam of which "paints," in a sense, an image of the

object, or scene, on the end of the tube. A television tube called the "Image Orthicon tube" shows much promise of making television a practical reality.

REVIEW QUESTIONS

1. What two types of radio transmitters are of the greatest importance?

2. Name some of the services for which radio transmitters are used.

3. What is meant by the term "speech-input equipment," and of what does it consist?

4. What is a preamplifier, a mixer, and a volume control? Why is a preamplifier often placed between the microphone and mixer, instead of after the mixer?

5. Explain the difference between low-level and high-level modulation.

6. If low-level modulation is used, what type of radio-frequency amplifier is employed?

7. What are parasitic oscillations, and how are they reduced?

8. What is overmodulation, what does it cause, and how can it be prevented?

9. Name several ways in which the power output of a radio transmitter can be measured.

10. Discuss the advantages and disadvantages of single-sideband transmission.

11. What two types of radio-telegraph transmitters are used? How does each operate?

12. What is a keying circuit? Describe one type.

13. Explain what is meant by the term "back wave," and explain what causes it.

14. What are key clicks?

15. Why is a lag circuit used?

16. Discuss briefly the methods of producing a frequency-modulated wave.

17. In what important ways do frequency-modulation transmitters differ?

18. Discuss one method of frequency stabilization used in a frequency-modulation transmitter.

19. Why are frequency doublers and triplers used in frequency-modulation transmitters?

20. What is meant by a subharmonic frequency, and how is it obtained?

21. What does a system of television accomplish?

22. What steps or processes are required in a television system?

23. What is meant by a television camera, and what is its function?

24. What type of modulation is used for transmitting the image signals in television? For transmitting the sound signals?

25. What would be the nature of the wave shape of the video signals if a television camera were scanning a black-and-white checkerboard?

PROBLEMS

1. The internal impedance of a certain dynamic microphone (page 29) is about 30 ohms at 1000 cycles. Should this microphone be connected directly to a 50-ohm studio-to-control-room cable circuit? Why? How would you recommend that it be connected to a 250-ohm circuit?

2. What will be the value of resistance of each resistor of the mixer of one channel of Fig. 244 when the mixer is to offer zero loss, 20-decibel loss, and an infinite loss in a 250-ohm impedance circuit?

3. When a certain radio transmitter is unmodulated, the output of the carrier is 1.0 kilowatt. High-level amplitude modulation is used, and the modulating audio amplifier has two tubes in class B. For 100 per cent modulation, what must be the approximate peak power-handling capacity of each tube?

4. An early single-sideband transoceanic radio-telephone transmitter radiated a 200-kilowatt signal when fully amplitude-modulated. What would have been the required capacity of the final amplifier if the carrier and other sideband had been radiated? What saving in power is obtained by using the single-sideband system?

5. Present amplitude-modulation transmitters must be held within plus or minus 20 cycles of the assigned carrier frequency. Express this in per cent for a 550-kilocycle station and for a 1600-kilocycle station. What device is used to maintain this frequency. For frequency modulation a station operating at 100 megacycles must hold the unmodulated carrier, or midfrequency, within plus or minus 2000 cycles. Express this as a percentage.

CHAPTER XIII

ANTENNAS AND RADIO TRANSMISSION

The purpose of a radio system is to transmit information, or programs, through space to a distant radio listener. The preceding chapters have discussed the nature of speech and program signals, and have explained how these are amplified and translated by modulation to radio frequencies for radiation into space by the transmitting antenna.

A **transmitting antenna** is a conductor, or system of conductors, for radiating radio waves into space. The antenna draws radio signal energy from the radio transmitter and radiates this signal energy. In a sense the antenna is an impedance-matching device between the transmitter (or transmission line or cable) and space. Without the antenna very little energy will be radiated into space.

Transmission Lines and Antennas.—The development of an antenna from a transmission line is of much fundamental importance. In Fig. 258*a* is shown an open-circuited transmission line one wavelength long. A connected radio transmitter is represented by the generator at the left end. The transmitter is assumed to be sending a pure sine wave of radio frequency. The distribution of the voltage along the line is shown by the solid lines, which is somewhat in accordance with the method followed in Chap. V. There is one important difference, however. The solid lines of Fig. 258*a* show the *magnitude of one-half the total voltage between wires*, that is, the voltage from each line wire to an imaginary reference plane midway between the wires. The magnitude (only) of the voltage at each point is indicated by the distance from the line wire to the solid voltage curve. The *direction* of the voltage *at a given instant* is indicated by *solid* arrows. Thus if an alternating voltage exists between two line wires (as it will at the open end and at other points) at a given instant, one wire will be positive and the other negative with respect to an imaginary zero plane midway between the wires.

The *magnitude* (only) of the current at each point in *each* line wire is shown by the distance between the line wire and the dotted

current curve. The *direction* of the current *at a given instant* is indicated by the dotted arrows.

Electric power is computed by the equation $P = EI \cos \theta$. Since in space (including air) the electric field E is directly propor-

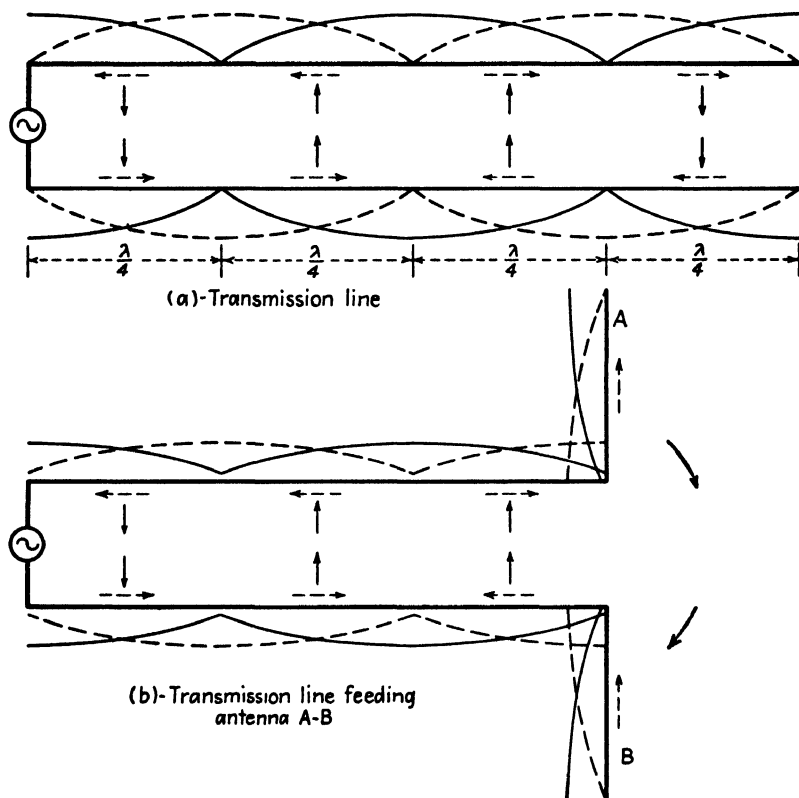



Fig. 258.—Showing how a transmission line is changed into a line driving an antenna by "bending back" a section one-fourth wavelength long. The voltage values (solid curves and arrows) are between the wires and an imaginary reference plane such as a conducting plane midway between the wires; or, the reference plane may be earth if the wires are symmetrical with respect to the surface of the earth.

tional to the voltage, and the magnetic field B is directly proportional to the current, it can be written that

$$P = k\text{EB} \cos \theta, \tag{124}$$

where θ is the *time* angle between the two fields, just as θ in the familiar power equation is the time angle of lead or lag and k is a constant. The equation just given is included because it indicates

that for power (or electromagnetic wave energy) to exist in space there must exist together at the same time an electric field E and a magnetic field B.

If Fig. 258*a* is examined, it is seen that the voltage arrows at a given point are in the same direction, and thus an electric field will exist between the wires. This field theoretically will extend out into space, as indicated in Fig. 68, page 125. A study of Fig. 68 will show, however, that at a point out in space the two magnetic-field components tend to cancel, because the two line currents are equal in magnitude and opposite in direction. Thus, according to Eq. 124, a transmission line will radiate little electromagnetic energy into space. Since the line of Fig. 258*a* radiates negligible energy, since it may be assumed lossless, and since it is open at the distant end, it cannot absorb power from the transmitter, and the input impedance will vary, as discussed on page 154.

Now suppose that the transmission line is opened out as indicated in Fig. 258*b*. The end parts *A* and *B* now form an antenna. The voltages between wire *A* and the reference plane and between wire *B* and the reference plane are in the same direction as before, and an electric field E will be established between antenna wires *A* and *B*. This electric field will extend out into space. Of great importance is the direction of the currents in wires *A* and *B* of Fig. 258*b*. Although these flow along the individual wires as before, at a given instant they are *in the same relative direction*, and *their magnetic fields do not cancel.* In fact, the magnetic fields add, so that in space a magnetic field B now exists.

It has been explained that antenna wires *A* and *B* produce a magnetic field B and an electric field E in space. If it can be explained that these fields are in *time* phase, or have components in time phase, then, according to Eq. (124), power will be radiated by the antenna. To explain this, reference is made to page 154. There it is shown that the input impedance of a quarter-wave section of a lossless line approaches a zero value of pure resistance. Now antenna wires *A* and *B* of Fig. 258*b* may be assumed lossless, but a transmitting antenna must radiate radio-frequency energy, or radio transmission would be impossible. Also, if an antenna radiates energy, it must draw energy from some source. The entire system of Fig. 258*b* may be regarded as a radio transmitter connected to a transmission line that in turn drives the antenna *A*-*B*. If *each* of the antenna wires *A*-*B* is one-fourth wavelength

long, the input impedance to the antenna *A-B* at the junction with the line will be a finite value of pure resistance. It will not be a zero value of pure resistance, because the antenna wires radiate power and must draw power. For the critical quarter-wave wire length specified, the reactance terms will be zero just as for a quarter-wave lossless line. Because the input imepdance of antenna wires *A-B* is pure resistance, the current that enters will be *in time phase* with the voltage, and in space, the magnetic and electric fields will rise and fall together, just as the current in, and voltage across, a resistor rise and fall together because they are in time phase. Thus power is taken by antenna wires *A-B*, and power is radiated.

In Fig. 258*a* it is assumed that the open-circuited transmission line is lossless and that no radiation occurs. Because of this the voltage and current standing waves reach a zero value. But in Fig. 258*b* the line is transmitting power to the antenna, and the standing waves do not rise to so high a value or fall to zero. Nevertheless, standing waves do exist along the antenna wires because the distant ends of the wires *A* and *B* are not terminated (except by "space"), and are, in a sense, "open." If it should happen that the input impedance (resistance when wire *A* and wire *B* are *each* one-fourth wavelength long) to the antenna equals the characteristic impedance of the connected transmission line (which is resistance at radio frequencies, page 139), then the line would be properly terminated, and no standing waves would exist along it. The general case of the antenna input impedance not equal to the line characteristic impedance has been assumed in Fig. 258*b*.

Induction and Radiation Fields.—For convenience, the antenna wires *A* and *B* of Fig. 258*b* have been reproduced in Fig. 259. The driving voltage from the transmitter is applied at the center. During the part of the voltage cycle that the upper generator terminal is *negative*, electrons will be repelled *up* wire *A*. At this same instant, the lower generator terminal will be *positive*, and electrons will be attracted *up* wire *B*. A conventional current will at this instant flow down in both wires *A* and *B* of Fig. 259, and a magnetic field will exist as indicated about the antenna. The deficiency of electrons in wire *B* will cause the lower end to be positive, and the excess of electrons in wire *A* will cause the upper end to be negative; by this action an electric field is established, as indicated in Fig. 259. One-half cycle later the relations will be

reversed. The representation in Fig. 259 is for one-half cycle later than in Fig. 258*b*.

As explained in the preceding section, the magnetic and electric fields will be in *time* phase and will rise and fall together. They are, however, at right angles, when viewed from the standpoint of their position *in space*. The field as a whole is expanding outward, and energy is being radiated outward as an electromagnetic wave. The direction of maximum radiation is along any line extending at

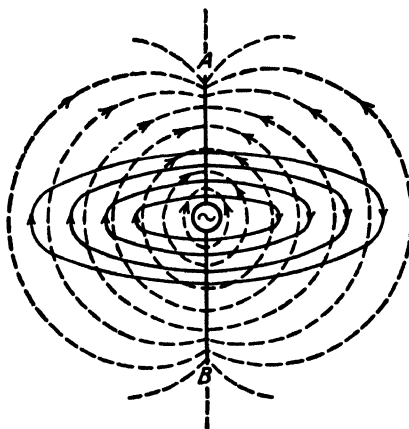



FIG. 259.—Showing the electric field (broken lines) and the magnetic field (solid lines) produced by the half-wave antenna *A-B* driven at the center. The next half cycle the directions of the lines will be reversed.

right angles from the center of the antenna in the plane in which the magnetic field is shown.

An important question may now be asked: Why does the energy in the fields not return to the circuit when the alternating voltage and current decrease? In elementary electrical theory it is always taught that when an alternating voltage builds up across a condenser, energy is stored in the condenser, and that when the voltage falls the energy is returned. Also, it is taught that when an alternating current builds up in a coil, energy is stored in the field of the coil, and that when the current decreases the energy is returned.

This confusing situation can be explained as follows: The magnetic field, produced by the current flowing in the antenna, and the electric field, produced by the difference in voltage between various parts of the antenna, combine to give an electromagnetic field about the antenna. This electromagnetic field consists of

two components. In the **near zone** close to the antenna an **induction field** exists. The strength of the induction field is inversely proportional to the square of the distance from the antenna. Energy *is* stored in the *induction field* when the voltage and current build up, and this energy *is returned to the circuit* when the voltage and current die out. This is the field usually considered in elementary electrical theory.

In the **far zone,** at a distance greater than, say, a few wavelengths from the antenna, the induction field approaches zero, and the **radiation field** exists. The radiation field contains the radio-frequency signal energy being radiated by the antenna. The strength of the radiation field at any distance away from the antenna is inversely proportional to the distance from the antenna. The energy in the radiation field does not return to the antenna when the antenna voltage and current die out. The energy in the radiation field becomes separated from the antenna. As a corresponding phenomenon, when a stone is dropped into the center of a quiet pond, long after the stone has settled on the bottom, water waves are traveling toward the shore, and the fact that these waves contain energy is indicated by the way in which the waves wash at the shore of the pond.

Antennas in Space.—Radio-transmitting antennas are of many types. The radiation from antennas is influenced by the presence of the earth's surface. The easiest way to study an antenna is to assume that it is isolated in space. This will be done in this section; later, the effect of the earth will be considered.

Each wire *A* and *B* of Fig. 258*b* was one-fourth wavelength long. Together, they constitute an antenna that is one-half wavelength long. Such an antenna is known as a **half-wave antenna** and sometimes is called a **dipole.** The radiation of this antenna *in space* will be considered now.

The shape of the field around a center-driven half-wave antenna is shown in Fig. 259. The question arises: In what direction, or directions, will it radiate the strongest radio signals? As has been explained, radio signals consist of electromagnetic waves, and these waves are composed of electric- and magnetic-field components. If an antenna radiates *only* one of these components in a given direction, then *no* radio energy is transmitted.

If the antenna of Fig 259 is examined, it is evident that the magnetic fields near the ends of the antenna wires are very small. The

ends of antenna wires *A* and *B* are open-circuited, and at these ends the current flow must approach zero. On the other hand, near the center of the antenna the current is large (Fig. 258*b*). Thus the magnetic field is strong in a plane at right angles at the *center* of the antenna. This is the plane shown in Fig. 259. Also, the magnetic field will be weak in a plane at right angles to the antenna near either end. The magnetic field in a direction along which the antenna points will be zero.

For the usual electric circuit, the electric field may be much stronger than the magnetic field. This is *not true* for the radiation field a few wavelengths from the antenna. In the radiation field, the magnetic and electric components of the electromagnetic wave contain equal amounts of energy.

It is of much value to know the strength of the electromagnetic wave that is radiated in each direction from an antenna. This information is necessary in the design of a radio system, because the signal energy should be so directed that the signal will be delivered to the geographical region desired. The directions of radiation from an antenna, and the magnitude of the radiation in each direction, are shown by graphical plots called "antenna patterns" or **radiation patterns.**

The *free-space* radiation pattern of the half-wave antenna of Figs. 258*b* and 259 is shown in Fig. 260. The vectors *A* and *A'* are the longest possible vectors from the antenna to the pattern boundary, and they indicate that maximum radiation occurs in these directions. Radiation in other directions is as indicated by the lengths of vectors, such as *B* and *B'*. The patterns are not quite circles. In studying Fig. 260*a*, it is *very* important to note that maximum radiation occurs in each direction, extending out at right angles from the center of the antenna. The pattern shown is that in one plane only, and there are an infinite number of planes. Thus the complete radiation figure for the antenna of Fig. 260*a* is a figure that would be generated (or created) if the pattern of Fig. 260*a* were revolved about the antenna as an axis. In other words, the complete antenna radiation is a "solid" figure, such as a "doughnut" with a small hole.

In Fig. 260*b* is shown the radiation pattern for the same half-wave antenna as viewed from the end. The radiation is equal in all directions in a plane at right angles to the antenna. It is assumed that the plane passes through the center; thus the diameter of the

pattern of Fig. 260*b* is the same as the length of vector A or A' of Fig. 260*a*. For some other plane at right angles to the antenna but *not* at the center of the antenna (for example, a plane near the end), the radiation pattern also will be a circle. The radius of the circle will not be so great, however, because the radiation near the ends of the antenna is small.

Free-space Radiation Patterns for Various Antennas.—The free-space radiation pattern of the half-wave antenna was considered in the preceding section. The discussion will now be extended to cover other antennas, using information from the references listed in the footnote.[1]

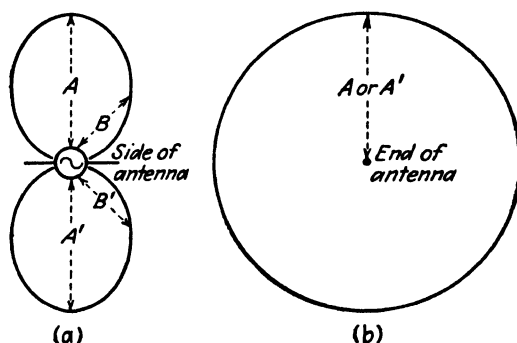



Fig. 260.—The radiation patterns of a half-wave antenna. (*a*) Radiation in a plane parallel to the direction of the antenna, and in which the antenna lies. (*b*) Radiation in a plane at right angles to the direction of the antenna, and passing through the center of the antenna.

When the total antenna length is an *odd number* of half wave-lengths (such as one-half wavelength), the radiated electric field is given by the relation

$$\epsilon = \frac{60I}{d} \frac{\cos\left(\pi \frac{L}{\lambda} \cos \theta\right)}{\sin \theta}, \tag{125}$$

and when the total antenna length is an *even number* of half wave-lengths (as when the antenna is one wavelength long)

$$\epsilon = \frac{60I}{d} \frac{\sin\left(\pi \frac{L}{\lambda} \cos \theta\right)}{\sin \theta}, \tag{126}$$

[1] See Terman, F. E., "Radio Engineers' Handbook," and P. S. Carter, C. W. Hansell, and N. E. Lindenblad, Development of Directive Transmitting Antennas by R. C. A. Communications, Inc., *Proceedings of the Institute of Radio Engineers*, Vol. 19, No. 10, October, 1931.

where ϵ is the electric-field strength in volts per meter, d is the distance from the antenna in meters, I is the maximum value in amperes of the current at any point in the antenna, L is the length of antenna in meters, λ is the signal wavelength in meters, and θ is the angle measured with respect to the antenna wire.

The first parts of these equations are constant terms. If they are neglected, and if the other terms are evaluated, the free-space radiation pattern results. This will be done for the half-wave antenna in which $L/\lambda = \frac{1}{2}$ in Eq. (125).

Case 1. $\theta = 30°$.

$$\frac{\cos(\pi/2 \cos 30°)}{\sin 30°} = \frac{\cos(1.57 \times 0.866)}{0.5} = \frac{\cos(1.36 \times 57.3°)}{0.5}$$
$$= \frac{0.208}{0.5} = 0.416$$

Case 2. $\theta = 60°$.

$$\frac{\cos(\pi/2 \cos 60°)}{\sin 60°} = \frac{\cos(1.57 \times 0.5)}{0.866} = \frac{\cos(0.786 \times 57.3°)}{0.866}$$
$$= \frac{0.707}{0.866} = 0.816$$

Case 3. $\theta = 90°$.

$$\frac{\cos(\pi/2 \cos 90°)}{\sin 90°} = \frac{\cos(1.57 \times 0)}{1.0} = \frac{\cos 0}{1.0} = \frac{1.0}{1.0} = 1.0$$

In evaluating these, the term $(\pi/2) \cos \theta$ will be in radians and must be multiplied by 57.3 to convert to degrees.

For a complete determination of the radiation pattern of a half-wave antenna, the radiation at 120°, 150°, etc., should be calculated. The pattern is symmetrical, however, and can be plotted from the data just calculated. This has been done in Fig. 261.

A plot of Eq. (126) will give the radiation pattern for a full-wave antenna. For this the angles should be taken at closer intervals, say every 10 or 15 degrees. When such data are calculated and plotted, Fig. 262 results.

Antennas Near the Surface of the Earth.—The radiation pattern from a center-driven half-wave antenna isolated *in space* was considered in the preceding section. Antennas often are close to the surface of the earth, and wave reflection from the earth has an important effect on the radiation pattern. It is assumed that the reflection occurs at the surface of the earth (rather than at some distance below the surface) and that the reflection is accomplished without loss at the surface.

A very convenient method of studying the effect of reflection is to consider that an antenna a short distance above the surface of the earth is accompanied by its "image" an equal distance below the surface. This is indicated in Fig. 263. It is important to note the polarities of the **image antennas**. These polarities are easily remembered because they are the polarities that the actual

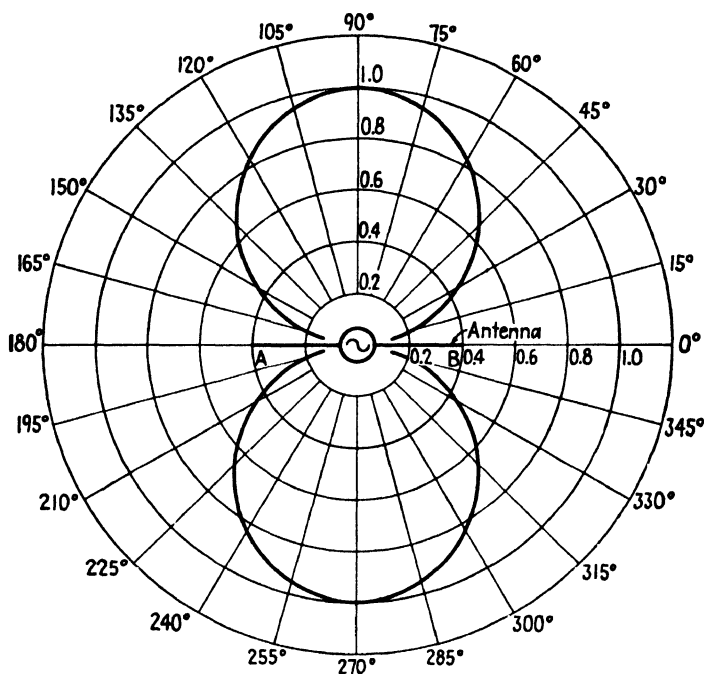



FIG. 261.—Plot of data from Eq. (125) giving the radiation pattern for the half-wave antenna *A-B*. The pattern is in a plane parallel to the antenna and in which the antenna lies.

antennas would induce in wires if wires were put in the place where, in effect, the image antennas are located.

Ground Reflection of Radiation from Horizontal Antennas.—In Fig. 264 is shown an end view of a half-wave antenna that is *horizontal* to the surface of the earth and at a height H above the surface. Electromagnetic wave energy constituting the radio signal to be transmitted is shown to be reaching a distant point by two paths: (*a*) as a direct ray and (*b*) as a reflected ray. The point at which the radiation is to be studied is assumed to be so far distant that the two rays are considered parallel. In determining the

strength of the signal that reaches a distant point, two factors must be considered: (*a*) the reflected ray that is *assumed* to be radiated by the image antenna when it travels a distance $2H \sin \theta$ greater than the direct ray; and (*b*) the fact that the radiation from the imaginary image antenna is 180° out of phase with that from the actual antenna, because of the way that the charges of Fig. 263 are distributed.

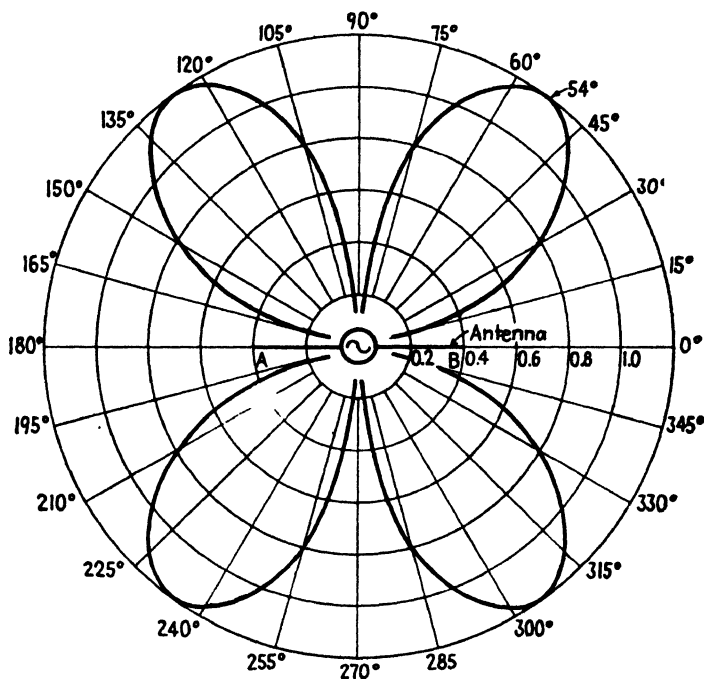



FIG. 262.—Plot of data from Eq. (126) giving the radiation pattern for the full-wave antenna *A-B*. The pattern is in a plane parallel to the antenna and in which the antenna lies.

The effect of reflection from the surface of the earth can be determined by studying the relative strengths of the combined signals that arrive direct from the antenna, and by reflection (that is, from the image). The method can be illustrated by a problem.

Illustrative Problem.—A horizontal antenna is located a distance of one-fourth wavelength above the surface of the earth. Compute the necessary data, and draw a diagram showing the effect of reflection from the earth.

Solution.—Step 1. Assume that the strength of the signal received via the *direct path* is 1.0. Assume that the phase angle of this received signal is zero (it is the reference vector). The angle by which the signal received via

the reflected path will lag the signal received via the direct path will be 180° + 180° sin θ. The first 180° is caused by the fact that the charges of the horizontal image antenna of Fig. 263 are as shown. The term 180° sin θ is because the wave from the image antenna must travel $2H$ sin θ farther, where H is one-fourth wavelength, or 90°. For various angles calculate the phase angle between the two vectors, and combine the vectors to find the resultant vector, which is the reflection factor to be plotted.

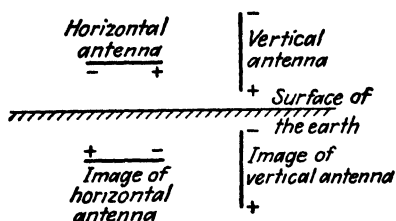



Fig. 263.—The surface of the earth reflects a wave sent to it from an antenna. The effect is as if an image antenna were located an equal distance below the surface of the earth. The polarities of the image antennas are as shown.

Step. 2. The radiation along the surface of the earth will be zero because the direct ray and the ray from the image each travel the same distance, and they are 180° out of phase.

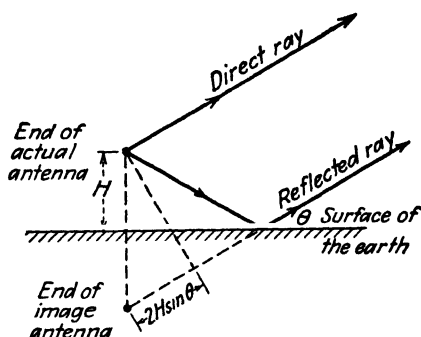



Fig. 264.—To reach a distant point, the reflected ray must travel a distance $2H$ sin θ farther than the direct ray. If vertical antennas are considered the actual and image antennas radiate in phase, but with horizontal antennas as shown here, the image antenna radiates 180° out of phase.

Step 3. Calculate the radiation at $\theta = 30°$. For this angle sin $\theta = 0.5$, and $180° + 180° \sin\theta = 180° + 180° \times 0.5 = 270°$. If it is assumed that the strength of the received direct ray is 1.0, and that it has zero angle, then the strength of the reflected ray which is considered to come from the image antenna also is 1.0, but it has an angle of lag of 270°. An angle of lag of 270° is equivalent to an angle of lead of 90°. Thus the two signals coming by the two paths can be represented by two vectors, each having a length of 1.0 and 90° apart. The resultant of two such vectors is $\sqrt{1^2 + 1^2} = \sqrt{2} = 1.41$. Thus, because of reflection, the radiation at an angle 30° above the horizon will have a strength 1.414 times as great as if no reflection existed.

Step 4. Calculate the radiation at $\theta = 60°$, at which sin 60 = 0.866. For this angle $180° + 180° \sin 60° = 180° + 180° \times 0.866 = 180° +$

156 = 336°. The reflected wave lags the direct wave by 336°, which is the same as leading it by 24°. Thus the received signal consists of two vector values, each of 1.0 magnitude and 24° apart. The received signal strength will be 1.0 × cos 12° + 1.0 cos 12° = 1.0 × 0.978 + 1.0 × 0.978 = 1.96.

Step 5. Calculate the radiation at θ = 90°, at which sin θ = 1.0. For this angle 180° + 180° sin 90° = 180° + 180° × 1.0 = 360°. The reflected ray (in this case being reflected straight up) will be in phase with the direct ray, and because it was assumed that no loss occurred at reflection, the radiation straight up is equal to twice that of the direct ray, or 2.0.

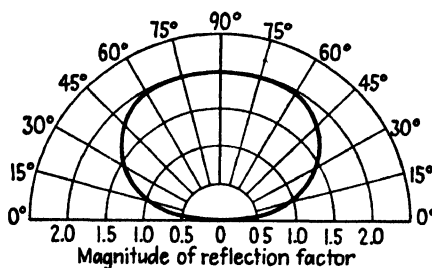



FIG. 265.—Effect of ground reflection on radiation from a horizontal antenna one-fourth wavelength above the surface of the earth. This is a plot of the reflection factors for radiation at the various angles. (See footnote, page 502.)

Step 6. Plot a diagram, showing the effect of reflection from the earth, for a *horizontal half-wave antenna one-fourth wavelength above the surface of the earth.* This has been done in Fig. 265. The preceding calculations are not entirely sufficient for the purpose, and for accurate plotting, calculations should be made at intermediate angles, such as θ = 15°, θ = 45°, etc. The left half of the diagram is the same as the right half.

Fig. 265 is a plot of the reflection factors by which the free-space radiation in various directions must be multipled to determine the actual radiation when a *horizontal antenna* is close to the earth. Of course horizontal antennas are at heights other than one-fourth wavelength above the surface of the earth. For this reason reflection diagrams other than Fig. 265 are needed. The diagram for a *horizontal half-wave antenna one-half wavelength above the earth* is shown in Fig. 266. This diagram is calculated in the same manner as the one explained in the preceding example. For a height of $H = \frac{1}{2}\lambda = 180°$, $2H \sin \theta = 2 \times 180° \times \sin \theta$. Thus, in the *vertical direction* the phase of the reflected wave would be the image out-of-phase angle of 180° + 360° sin θ = 180° + 360° × 1.0 = 540°. An angle of lag of 540° is equivalent in effect to an

angle of lag of 180°. Thus, at points directly above the antenna the direct and reflected waves cancel, as Fig. 266 indicates.[1]

Ground Reflection of Radiation from Vertical Antennas.—Horizontal half-wave antennas were considered in the preceding section. In this section, reflection from a *vertical half-wave antenna* at various heights above the earth will be discussed. The vertical antenna and its image are shown in Fig. 263. It will be noted that the two are *in phase.* Figure 264 may be used for studying the strengths of the received signal. The two rays are treated as coming from the center of the antenna and from its

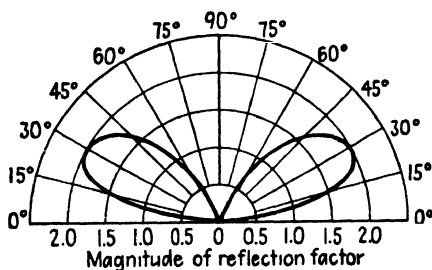



FIG. 266.—Effect of ground reflection on radiation from a horizontal antenna one-half wavelength above the surface of the earth. This is a plot of the reflection factors for radiation at the various angles. (See footnote.)

image, and height H is the distance from the *center* of the antenna to the surface of the earth.

The reflection-factor diagram for a vertical half-wave antenna with the center one-fourth wavelength above the surface of the earth is shown in Fig. 267, and it is arrived at as follows: At zero degrees (along the horizontal) the direct ray and the reflected ray (assumed to come from the image antenna) both travel the same distance to reach a distant point, and because the instantaneous antenna phase relations are the same (Fig. 263), the two waves arrive in phase. If the radiation of the direct ray is 1.0, and if no loss occurs at reflection, then the reflected ray also will be 1.0, and the reflection factor will be 2.0. At an angle of 30°, the reflected ray will travel a distance of $2H \sin \theta$ farther than the direct ray. Where H is the height above the earth in wavelengths (one wavelength equals 360°), for $H = \frac{1}{4}\lambda$, the angle by which the re-

[1] The reflection-factor diagrams of Figs. 265, 266, 267, and 268 are shown here essenitally as they appear in "The A.R.R.L. Antenna Book," a publication of the American Radio Relay League. This book contains diagrams for other heights such as $H = \frac{1}{8}\lambda$, $H = \frac{3}{8}\lambda$, etc., and also contains much additional information on antennas.

flected ray lags the direct ray at any distant point is $2 \times 90° \times \sin 30° = 180° \sin 30° = 180° \times 0.5 = 90°$. Thus the received signal at a distant point will consist of two vectors, each of magnitude 1.0 and 90° apart in time phase. Hence, the strength of the combined signal at an angle of 30° with the horizontal is 1.414. For $\theta = 90°$, the relations are $2H \sin \theta = 180° \times \sin 90° = 180° \times 1.0 = 180°$.

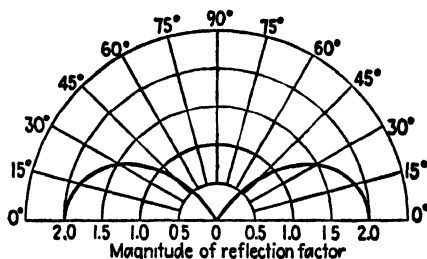



FIG. 267.—Effect of ground reflection on radiation from a vertical antenna one-half wavelength long, and with its center one-fourth wavelength above the surface of the earth. This is a plot of the reflection factors for radiation at the various angles. (See footnote, page 502.)

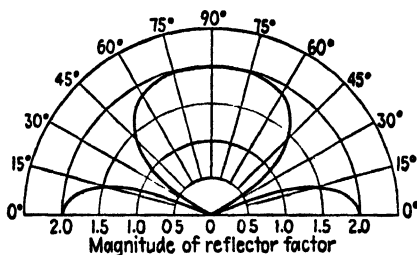



FIG. 268.—Effects of ground reflection on radiation from a vertical antenna one-half wavelength long, and with its center one-half wavelength above the surface of the earth. This is a plot of the reflection factors for radiation at the various angles. (See footnote, page 502.)

Therefore, the reflected signal is 180° out of phase with the direct signal, and cancellation occurs along a vertical line.

The reflection factors for a *vertical half-wave antenna with its center one-half wavelength* above the surface of the earth are shown in Fig. 268. At zero degrees, the two rays travel the same distance, and because they are radiated in phase (Fig. 263), they add directly at a distant point, giving a factor of 2.0 along the horizontal. At 30°, the reflected ray that is assumed to come from the image antenna has to travel a distance $2H \sin \theta$ or $2 \times 180° \times \sin 30° = 360° \times 0.5 = 180°$ farther than the direct ray, and will

arrive 180° out of phase with it. Cancellation occurs, and the reflection factor at 30° is zero, as Fig. 268 indicates. At the vertical, or 90° angle, $2H \sin \theta = 360 \times 1 = 360$. Thus the two rays are in phase, and the reflection factor is 2.0. Of course vertical antennas have heights other than $\frac{1}{4}\lambda$ and $\frac{1}{2}\lambda$, the values considered in this section. Diagrams for such heights can be found in the reference listed in the footnote on page 502.

Radiation Patterns for Horizontal Half-wave Antennas.—The fact is stressed again that the diagrams of Figs. 265, 266, 267, and 268 *are not* radiation patterns. Instead, they are plots of the factors by which the free-space radiation patterns must be modified to take into account the effect of reflection from the surface of the earth. With the free-space pattern known and the reflection factors available, the actual radiation pattern for an antenna close to the earth can be computed.

To prevent confusion, great care must be used in describing exactly what a radiation pattern represents. Such patterns indicate the relative strengths of the radiation at various angles *in a given plane.* Just where this plane is situated with respect to the antenna is *very* important.

Vertical-plane Radiation from Horizontal Half-wave Antennas.—The first case which will be considered is the radiation pattern in the *vertical plane is which the antenna lies*, the antenna being *one-fourth wavelength* above the surface of the earth. In Fig. 269, this is seen to be the vertical plane in the direction of the antenna wires *A-B*. If the antenna were in free space, the radiation would be as shown by the dotted line. For convenience, the maximum free-space radiation (at right angles to *A-B*) will be considered as *unit radiation.* To find the *actual* radiation, the reflection effect of the earth must be considered. At an angle of 20° with the horizontal, the direct, or free-space, radiation is $R_1 = 0.28$. (This is not shown in Fig. 269.) From Fig. 265, the 20° ground reflection factor is about 1.1. Thus the actual radiation at 20° is $0.28 \times 1.1 = 0.31$. This is plotted on the 20° line. The free-space radiation at 40° is $R_2 = 0.66$. From Fig. 265, the 40° ground reflection factor is 1.7. The actual radiation at 40° is $1.7 \times 0.66 = 1.12$, which is plotted on the 40° line. The free-space radiation at 60° is $R_3 = 0.81$, and the corresponding reflection factor is 1.95. The actual radiation is $0.81 \times 1.95 = 1.57$, and this is plotted at 60°. The direct radiation at 80° is about 0.97, and the reflection factor

is 2.0, giving an actual radiation of 1.94. When these data are plotted, the solid curve of Fig. 269 results. The distance from the center of the antenna to the (solid) curve at a given angle represents the relative radiation at that angle compared with the maxi-

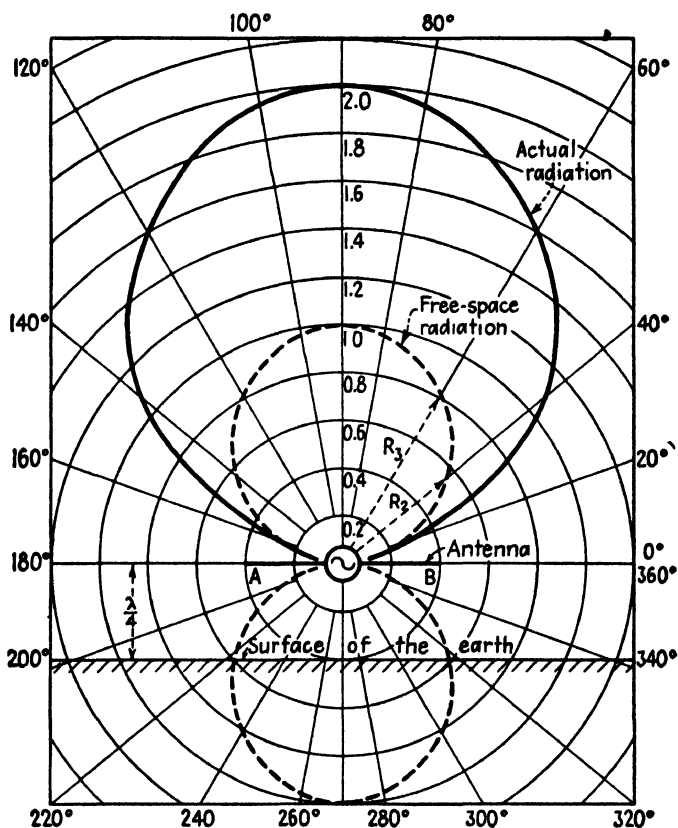



FIG. 269.—The dotted figure is the free-space radiation figure for the half-wave antenna A-B. The plane is parallel to the antenna and is the plane in which the antenna lies. The plane is vertical to the earth. The solid figure shows the actual radiation pattern if the antenna is *horizontal* and *one-fourth wavelength* above the surface of the earth. The relative radiation at each angle is the length of a line from the center of the antenna to the solid line. The areas of the figures should be disregarded.

mum radiation in the *vertical* direction. The maximum radiation is 2.0 for convenience only; the numerical values have no other meaning.

The next case that will be considered is the radiation pattern in a *vertical plane passing through the center and at right angles* to

the horizontal half-wave antenna *one-fourth wavelength* above the surface of the earth. As in the preceding paragraph, the free-space

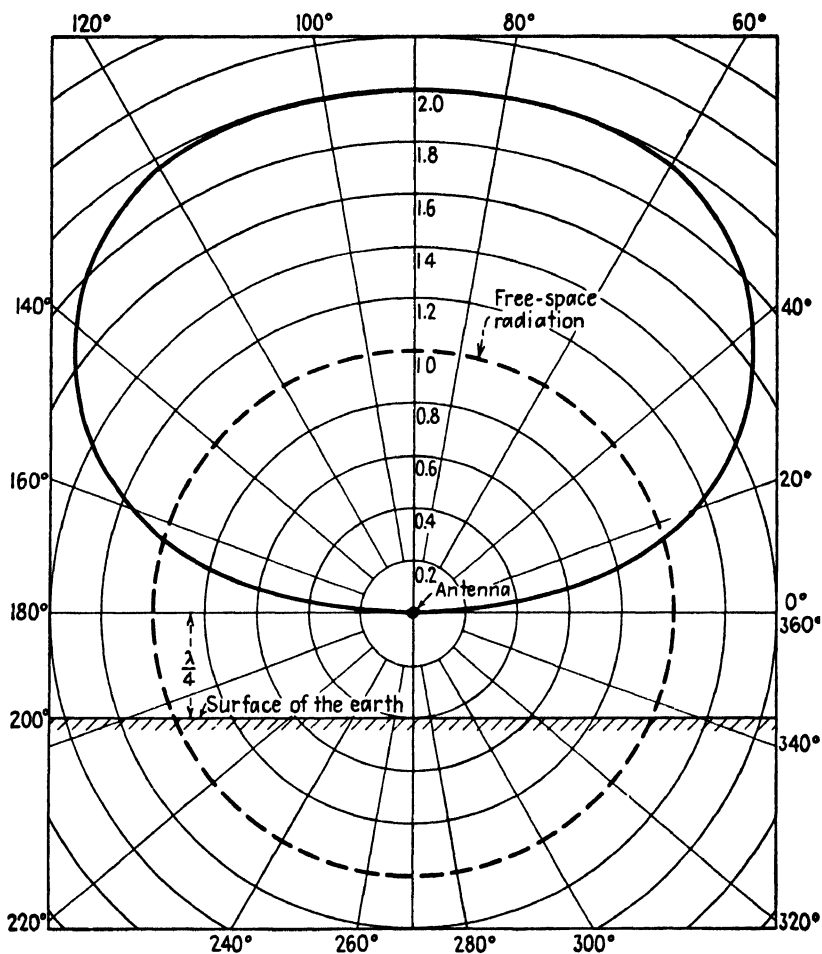



Fig. 270.—The dotted figure is the free-space radiation pattern from a half-wave antenna in a plane at right angles to the direction of the antenna and passing through the center of the antenna. The plane is vertical to the earth. The solid figure shows the actual radiation pattern if the antenna is *horizontal* and *one-fourth wavelength* above the surface of the earth. The relative radiation at each angle is the length of a line from the antenna at the center to the solid figure. The areas of the figures should be disregarded.

radiation pattern is corrected by the ground-reflection factors in order to determine the actual radiation pattern as modified by reflection from the surface of the earth. The position of the antenna

wire is shown in Fig. 270, the antenna being at right angles to the plane of the paper. The dotted circle shows the shape of the free-space radiation pattern. This must be modified by the reflection factors obtained from Fig. 265. Since the free-space pattern is unity in each direction, the actual radiation pattern will have the same shape as the ground-reflection factor diagram, and is as shown by the solid line of Fig. 270.

The antenna just considered was one-half wavelength long, horizontal, and one-fourth wavelength above the surface of the

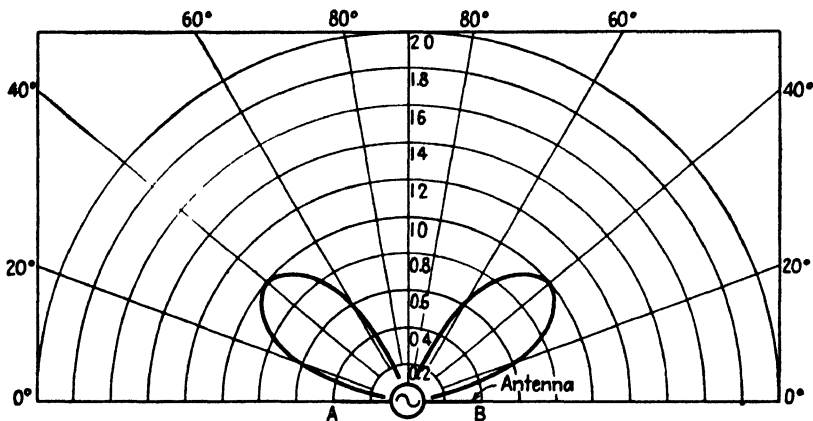



FIG. 271.—Radiation pattern for the horizontal half-wave antenna *A-B one-half wavelength* above the surface of the earth. The plane is vertical to the surface of the earth, parallel to the antenna, and is the plane in which the antenna lies.

earth. The following discussions will be for the same antenna *one-half wavelength* above the surface of the earth. The free-space pattern in a *vertical plane in the direction of the antenna and in which the antenna lies* is shown by the dotted pattern of Fig. 269. This must be corrected for ground reflection, using the factors from Fig. 266. The method is the same as for determining Fig. 269, except that the reflection factors from Fig. 266 instead of Fig. 265 are used. The calculations give the curve of Fig. 271.

The radiation pattern in a *vertical plane passing through the center and at right angles* to a horizontal half-wave antenna *one-half wavelength* above the surface of the earth will be considered. The free-space pattern is a circle, as shown in Fig. 270. This must be modified by the reflection factors of Fig. 266. Because the space-free pattern is a circle, the actual radiation pattern will be as shown in Fig. 272.

It is again stressed that these patterns should be interpreted merely as showing the relative radiation in the various directions with respect to the horizontal.

Horizontal-plane Radiation from Horizontal Half-wave Antennas. —Radiation patterns in *vertical* planes were discussed in the preceding paragraphs. If the radio waves were visible, and if the region about the antenna could be examined plane by plane, these patterns would be seen by an observer standing on the surface of

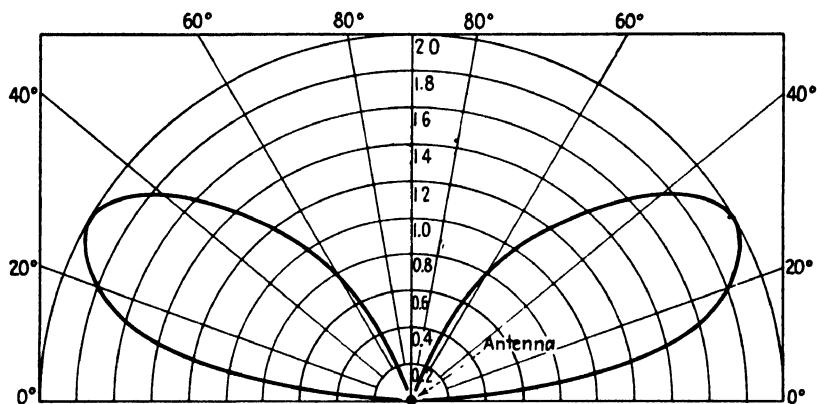



FIG. 272.—Radiation pattern for a horizontal half-wave antenna *one-half wavelength* above the surface of the earth. The plane is vertical to the surface of the earth, at right angles to the antenna, and passes through the center of the antenna.

the earth and looking toward the antenna, first at the side, and next at the end. The discussion immediately following considers the radiation in *horizontal* planes. These are the radiation patterns that an observer would "see" if he were directly above the antenna and "looking down" and examining the radiation from the antenna in the *various* horizontal planes. The horizontal patterns depend on where the plane is situated with respect to the antenna. For instance, Figs. 269 and 270 show that theoretically a horizontal half-wave antenna one-fourth wavelength above the surface of the earth radiates no signal in the various horizontal directions in which the antenna lies. But, if the radiation is studied in some other plane, there is radiation and a pattern. For instance, on a plane where the two patterns intersect the 30° angle line, there is a radiation pattern that can be determined readily. For example, the radiation at 30° *in the direction of the antenna* is about 0.75 units, according to Fig. 269. The horizontal radiation would be 0.75

cos 30° = 0.75 × 0.866 = 0.65, approximately. Also, the radiation at 30° at right angles to the antenna is 1.7 from Fig. 270. The horizontal radiation in a plane passing through this point is 1.7 × 0.866 = 1.5, approximately. Thus the radiation pattern in a

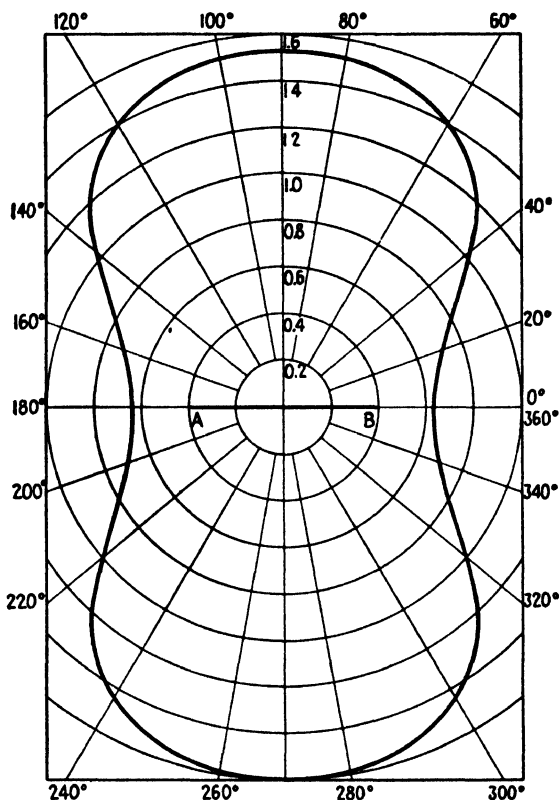



FIG. 273.—Radiation pattern for a horizontal half-wave antenna that is one-quarter wavelength above the surface of the earth. The pattern is in a horizontal plane parallel to the antenna, but at an angle of 30° (see text). This is a view "looking down" on the antenna and its pattern in the 30° horizontal plane.

horizontal plane at the 30° level is as shown in Fig. 273. Of course plots could be made for planes at 40°, etc.

Radiation Patterns for Vertical Half-wave Antennas.—The radiation patterns for *vertical* half-wave antennas at several heights above the surface of the earth will be developed in this section. As before, the free-space pattern is corrected for reflection from the earth, and this gives the actual radiation pattern.

Vertical-plane Radiation from Vertical Half-wave Antennas.—The first case to be considered is the *vertical-plane* radiation from a vertical half-wave antenna with its center *one-fourth wavelength* above the surface of the earth. The antenna lies in the plane. The free-space pattern is shown dotted in Fig. 274. This is modified by the ground-reflection chart of Fig. 267. At 0° with the horizontal, the free-space radiation is 1.0, and from Fig. 267 the reflection factor is 2.0, giving a radiation along the earth of 2.0.

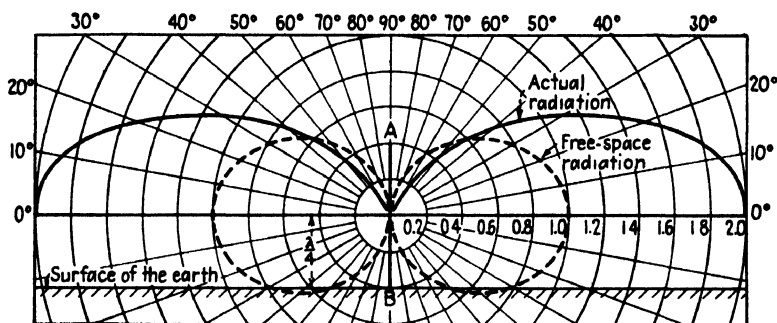



FIG. 274.—The half-wave antenna *A-B* in free space has a radiation pattern as shown by the broken line. When this antenna is *vertical* and *one-fourth wavelength* above the surface of the earth, reflection from the earth modifies the radiation so that the actual pattern is as shown by the solid lines. This represents the radiation in a vertical plane in which the antenna lies.

At 10° with the horizontal, the space-free radiation is 0.97 unit, and from the ground-reflection chart the reflection factor is 1.9, giving an actual radiation of 1.84, which is plotted (solid line) on Fig. 274. At 20°, the free-space radiation is 0.9, the reflection factor is 1.7, and the actual radiation is 1.53, which also is plotted on Fig. 274. At 30°, the actual radiation is $0.8 \times 1.4 = 1.12$, at 40°, it is $0.7 \times 1.1 = 0.77$, at 50°, it is $0.56 \times 0.7 = 0.39$, and at 60°, the actual radiation is $0.4 \times 0.5 = 0.2$ unit. The plot of these computed actual radiation values is, as mentioned, the solid line of Fig. 274. This shows very little signal radiation in the upward direction, and much radiation along the surface of the earth. The signal along the surface would not be so strong as this pattern indicates, because it assumes that the surface is lossless. Actually, there is considerable loss in the surface of the earth, and also in objects on the surface, such as buildings, trees, etc., particularly at high frequencies. Since the vertical antenna is symmetrical, one

diagram is sufficient to show the radiation in any vertical plane in which the antenna lies.

The center of the antenna just considered was one-fourth wavelength above the surface of the earth, and the radiation pattern found was for a vertical plane in which the antenna lies. The vertical half-wave antenna will now be considered when its center is one-half wavelength above the surface of the earth. The

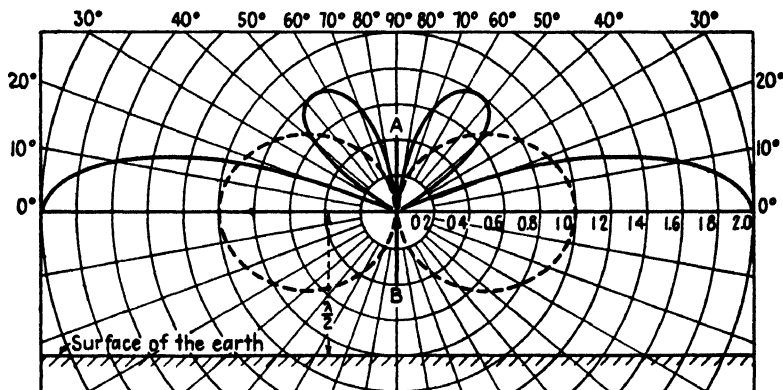



Fig. 275.—The radiation pattern of the half-wave antenna *A-B* in free space is shown by the broken line. When this antenna is *vertical* and *one-half wavelength* above the surface of the earth, reflection modifies the pattern so that the actual radiation is as shown by the solid lines. Note the formation of lobes at about 55°. This pattern represents radiation in a vertical plane in which the antenna lies.

method of calculation is the same as for Fig. 274. The space-free radiation pattern is corrected by factors from the ground-reflection chart of Fig. 268. This gives the pattern of Fig. 275, and it will be noted that lobes or "ears" have developed at about 55°. This means that much signal energy will be directed upward at this angle.

Horizontal-plane Radiation from Vertical Half-wave Antennas.—The preceding discussion was for radiation in the vertical plane, and if radiation were visible; it is the radiation that an observer would see if he were on the surface of the earth and looking toward the antenna. The *horizontal-plane radiation pattern* is to be discussed now. It is the pattern that an observer would see if he were directly above the vertical antenna and looking down on it. The vertical antenna is perfectly symmetrical, and the radiation pattern in a horizontal plane is a circle, as indicated in Fig. 276.

The diameter of the pattern for radiation in a horizontal plane on the surface of the earth will be greater than for a plane at right angles to the antenna but above the surface of the earth. Also, the radiation in the various planes may be different for the vertical half-wave antenna when its center is one-fourth wavelength above

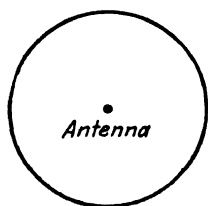



FIG. 276.—The radiation pattern in a horizontal plane for a vertical antenna is circular, indicating equal radiation in all directions along the surface of the earth.

the surface of the earth, and when its center is one-half wavelength above the surface. These facts can be verified by examining Figs. 274 and 275, and by imagining planes passing horizontally through

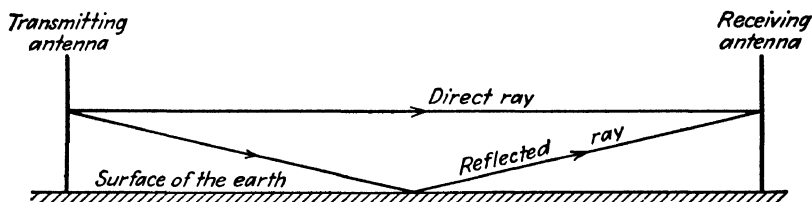



FIG. 277.—The space wave arriving at the receiving antenna is part of the ground wave. The space wave is composed of the two parts shown, the direct ray and the ground-reflected ray.

these patterns at various heights. Attention is again called to the fact that these figures assume that the surface of the earth is lossless.

Radio-wave Propagation.—A radio signal composed of electromagnetic waves in traveling from a transmitting antenna to a receiving antenna may follow one or more paths. Part of the energy may travel near the surface of the earth, and part may travel by a path that often is remote from the surface of the earth. These possible paths will be considered now.

The Ground Wave.—The component of the radio wave that is *affected by* the surface of the earth is termed the **ground wave.** The ground wave is composed of two components, (*a*) a **surface wave** and (*b*) a **space wave.** The surface wave travels entirely along the surface of the earth. The space wave is the resultant of two rays, the **direct ray** and the **ground-reflected ray** (Fig. 277).

There are many factors determining the extent to which the surface wave and the ground-reflected wave are affected by the surface of the earth. The resistivity and dielectric constant of the surface layers are important factors because the wave penetrates the surface to an extent determined in part by the frequency of the signal being transmitted. Also, the heights and types of antennas affect the surface losses. Other factors are the curvature of the earth and the condition of the lower atmosphere.

The Sky Wave.—The ground wave accounts for part of the radio-signal energy that reaches a *distant* receiving antenna; the sky wave

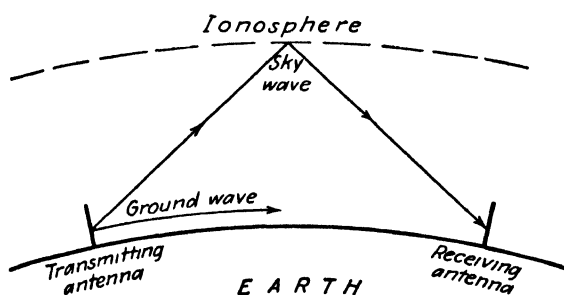



FIG. 278.—The sky-wave component radiated by a transmitting antenna is reflected back to the surface of the earth by the ionosphere.

accounts for the remainder. The **sky wave** consists of signal energy that is radiated upward by the transmitting antenna, and reflected back toward the earth. As Fig. 272 indicates, for certain antennas much energy is radiated upward at an angle of 30°. If this is reflected back to the surface of the earth, then radio reception is possible by the sky wave alone. If the ground wave also is delivering energy to the receiving antenna, then the magnitude of the total received signal will be the vector sum of the signals received by the two paths. This is indicated by Fig. 278. The sky wave may bounce back and forth several times before it is attenuated to a negligible level.

The Ionosphere.—As is indicated in Fig. 278, the sky-wave component radiated by the transmitting antenna is reflected back to the earth by the ionosphere. If it were not for this reflecting medium, the sky wave would not return to earth, which of course means that communication by the sky-wave path alone would be impossible.

The ionosphere is a region of ionized gas that exists in the rarified

upper atmosphere of the earth. For example, at 60 miles above the earth the air pressure is about 1/1,000,000 of that at the surface, and the atmosphere largely is nitrogen. At higher levels, the air pressure is decreased further, and the content becomes largely helium. These air pressures are what would be called good vacuums. These rarified layers of gas are ionized largely by ultraviolet radiation from the sun. Ionized gas contains positive and negative charges, and is therefore electrically conducting. When electromagnetic waves used in radio are directed toward the ionosphere, the waves are bent back toward the earth, provided that

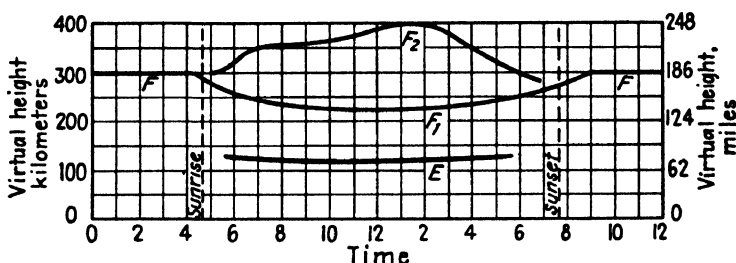



Fig. 279.—Reflecting ionized layers in the ionosphere for a typical 24-hour period in June. Actually the *E* layer exists to some extent during the night period. (Data from reference listed in footnote, page 517.)

their frequency is not too great. Very short high-frequency waves pass through the ionosphere and out into space.

The ionized layers of the ionosphere are far from constant. They vary during the 24-hour period, and they also vary with the season of the year, and in other ways. Two of the layers are more or less permanent, and two are semipermanent. The two permanent ionized layers are termed the **E layer** and the **F layer.** The *E* layer exists both day and night at various intensities and at a height of from about 55 to 85 miles. The *F* layer exists as a single layer at night at heights of from 110 to 250 miles. In the daytime the *F* layer divides into the **F_1 layer** and **F_2 layer.** The height of the F_1 layer is 85 to 155 miles. On a summer day the F_2 layer is at 155 to 220 miles, and on a winter day the F_2 layer is at 90 to 185 miles. A **D layer** has been found to exist in the daytime below the *E* layer, but the *D* layer apparently is of little importance. A typical arrangement for the various layers is as shown in Fig. 279.

Unfortunately the ionosphere does not reflect all the energy

striking it. Apparently much is absorbed by the charged ions that are affected by the radio wave. Little is known about this absorption except that it varies in degree, and that it is of much importance in causing attenuation. Absorption occurs because the radio waves penetrate the ionized layer to some extent before they are reflected. Thus reflection is not from the layer *surface*, but from a "virtual layer" having a "virtual height" above the surface of the earth.

Critical Frequencies and Critical Angles.—Reflection of radio waves from the various ionized layers in the ionosphere is caused by a bending, or **refraction,** as the wave travels in the ionized atmosphere. Theories have been advanced that the charged ions absorb energy from the radio signal and reradiate the energy in such way that the direction of travel of the wave front is bent back toward the earth. In any event, the wave enters the ionized layer, and its direction of travel is changed gradually.

There is a limit to this, however. If the frequency of the signal wave is increased (which means that the wavelength is decreased), a point is reached finally where the wave, although it may be bent somewhat, does not return to the earth, but at least some of the energy passes through the layer and out into space. The rest of the wave energy would be absorbed in the layer.

The height, or more properly the apparent or **virtual height,** of a layer is determined by sending upward a series of short pulses of energy of radio frequency and measuring the time taken for these pulses to travel to the reflecting layer and back to the earth. It will be recognized by many readers that these measurements, which have been in use for many years, utilize the basic principle of a radar system.

If a pulse of a single-frequency radio energy is being sent *vertically* upward, and if the frequency is increased, a critical frequency will be found at which no energy is returned to the earth. This is indicated by ray A of Fig. 280. If a directional antenna is used, and if the angle at which the wave is sent up is decreased, as shown at B and C, when the **critical angle** θ of Fig. 280 is reached, the wave *is* reflected to the earth. Also, reflection occurs, as indicated by rays D, E, and F, for angles smaller than the critical angle. There is accordingly a **skip distance,** as indicated in Fig. 280, and in this region no reception by sky waves is possible for

the particular frequency and layer height illustrated in Fig. 280. Of course a different frequency might put a signal into the skip-distance region of Fig. 280.

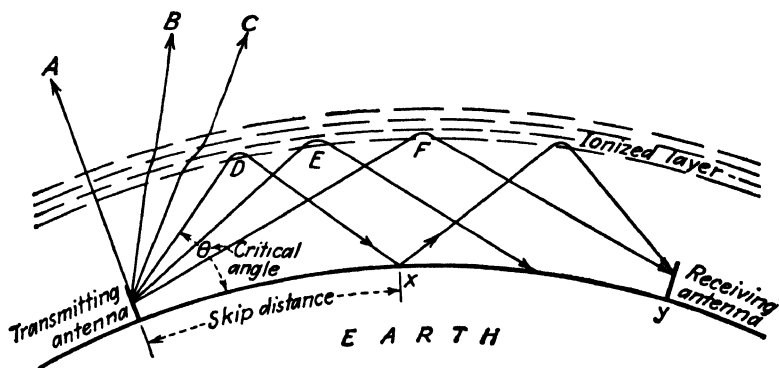



FIG. 280.—Showing how sky-wave signal energy may arrive at a distant receiving antenna by a "one-hop" or a "two-hop" path. Rays A, B, and C are not reflected back to the earth because they are directed upward at an angle greater than the critical angle. This figure shows the effect of varying the angle of radiation.

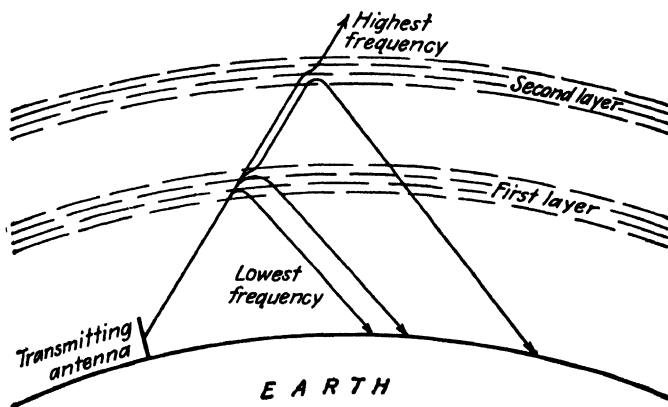



FIG. 281.—Showing the effect on reflection of varying the frequency. A wave of a given frequency might penetrate one layer and be reflected by a layer above.

This figure shows another phenomenon, that of radio communication by multiple reflection. Thus, a radio-receiving set located at point x would receive a signal sent out by the transmitter. If the sending antenna radiated only at the critical angle and sent only ray D, then a receiver located at point y would receive a signal after a second reflection from an ionosphere layer. If the trans-

mitting antenna sent out signals only at the angles at which D and F are shown, then the radio-receiving set would receive signals by two paths, a "one-hop path" and a "two-hop path."

The effect of varying the angle of radiation was shown in Fig. 280. A critical angle was said to exist at which energy ceased to be reflected back to the earth. On the other hand, reflection might occur from a higher ionized layer, as Fig. 281 indicates. Thus if the angle of radiation is held constant but the frequency is increased, a frequency might be found at which reflection from a low layer does not exist, but does occur from a higher layer. As the

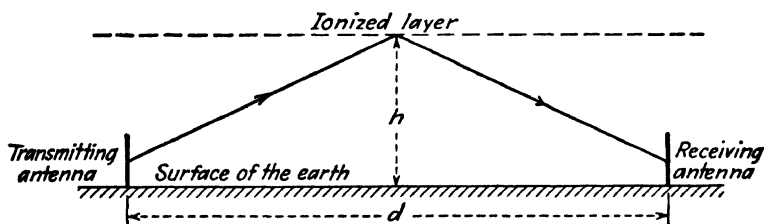



FIG. 282.—Showing dimensions to be used with Eq. (127) for finding the maximum usable frequency.

frequency is increased still further, even the higher layers cause insufficient bending to force the ray back to earth.

It is sometimes of interest to know the maximum frequency that can be used to communicate between two points; this is termed the **maximum usable frequency,** and is given by the relation[1]

$$f_m = f_c \sqrt{\frac{d^2}{4h^2} + 1}, \qquad (127)$$

where d is the horizontal distance between the two points, h is the virtual height of the reflecting layer, and f_c is the critical frequency, as previously defined. These relations are illustrated in Fig. 282. The distances h and d should be in the same units. This equation neglects the curvature of the earth.

Practical Aspects of Radio Transmission.—Certain of the theoretical aspects of radio transmission have been discussed in the preceding pages. In presenting the practical aspects of the sub-

[1] This equation and Fig. 282 on which it is based are from an article by J. H. Dellinger, The Role of the Ionosphere in Radio-Wave Propagation, *Transactions of the American Institute of Electrical Engineers*, Supplement, Vol. 58, p. 803, 1939. This article is a very complete treatment of the subject, and is recommended highly.

ject, there are two methods of transmission to consider, (*a*) by ground waves and (*b*) by sky waves.

Ground-wave Transmission.—The ground wave is the component of the radiation from an antenna which travels *along* the surface of the earth and which accounts for all the signal strength at a receiving antenna if the sky wave is excluded. The ground wave *does not* travel up to the ionosphere and back to the earth. The ground wave was explained on page 512 to be composed of a surface wave and a sky wave. The surface wave travels *entirely* along the surface of the earth, and is much affected by the nature of this surface. When both the transmitting antenna and the receiving antenna are at the surface of the earth, the surface wave is the entire ground wave for all but the very high frequencies. The surface wave is of great importance because in radio broadcast the transmitting and receiving antennas are at the surface of the earth. In fact, during the *daytime* reception of conventional amplitude-modulation radio-broadcast programs, the signals arrive at the receiving station by the surface-wave path. Antennas at the surface of the earth and a ground wave composed entirely of surface wave will be the only condition considered in the paragraphs that follow.

The ground wave radiated by a vertical antenna spreads out and the field strength should vary, theoretically, inversely as the distance. Such a decrease is shown by the inverse-distance curve of Fig. 283. Actually, the losses in the surface of the earth cause the curves at various frequencies to decrease as indicated. The passing radio waves cause currents to flow in the surface of the earth, and since the earth is not a perfect conductor, energy is absorbed. In general, broken mountainous terrain shows low conductivity, and flat terrain shows high conductivity. Of course the chemical nature of the soil, moisture content, and other factors are important.

The field strengths at the various locations are measured in **microvolts, or millivolts, per meter.** Theoretically a field strength of 1 microvolt per meter will induce a voltage of 1 microvolt in a wire 1 meter long if the wire is parallel to the electric lines of force in the oncoming field. The ground wave sent out by a grounded vertical broadcast antenna of the usual type is *vertically polarized*; that is, the electric lines of force are vertical to the surface of the earth (see Fig. 259). They must be vertical, because the earth is a conductor, and electric lines of force parallel to the earth would

be "shorted out." The electric field parallel to and *at the surface* of a good conductor approaches zero, because it causes the flow of

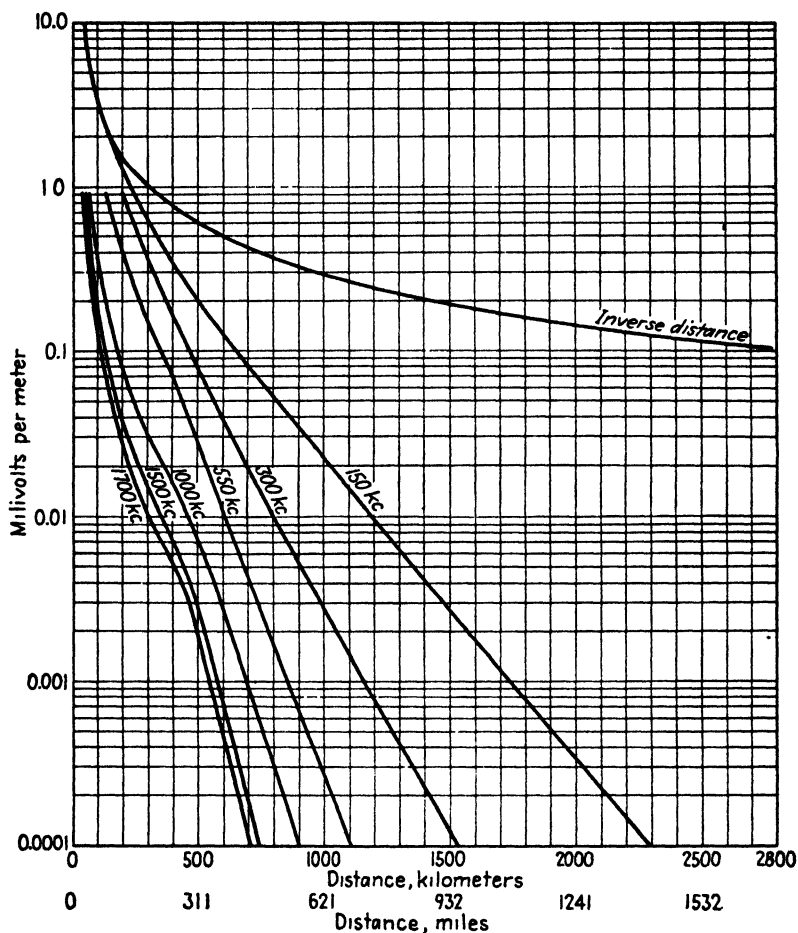



FIG. 283.—Showing how the field strength in millivolts per meter (page 518) varies with distance for a vertical antenna radiating 1.0 kilowatt of power at various frequencies. The ordinary amplitude-modulation band is from about 550 to 1700 kilocycles. For other values of radiated power the field intensities are proportional to the square root of the power. These curves are for the ground wave. The loss is largely due to poor ground conductivity. These curves are for a conductivity of $\sigma = 10^{-13}$ (c.g.s. electromagnetic units). (These curves are from *Proceedings of the Institute of Electrical Engineers*, Vol. 21, No. 10, October, 1933.)

large currents in the conductor and these absorb the energy in the electric field.

The strength of the field required for good broadcast-program reception depends on the locality. At a location that is free from interference a field strength as low as 10 microvolts per meter may give satisfactory results. In a location where there is much electrical interference that will be picked up by the radio receiver in addition to the desired radio signal, the required field strength may be 100 millivolts per meter.

As the frequency of a radio signal is increased, the losses encountered by the ground wave increase rapidly. This is indicated by Fig. 283. The ground wave is particularly useful at the very low frequencies and in the ordinary broadcast band, which is from about 550 to 1700 kilocycles. At much higher frequencies, the ground wave is absorbed so rapidly that it is useful only for a few miles from the transmitting antenna.

The ground wave is especially useful, because signals received via the ground-wave path are very constant and reliable as compared with those arriving by the sky wave. On the other hand, the distances that can be covered are limited, unless very large amounts of power are put into the transmitting antenna.

Sky-wave Transmission.—Radio signals arriving at a receiving station by sky-wave transmission have experienced one or more reflections from one of the several ionosphere layers; also they have been reflected from the surface of the earth if more than one "hop" has been involved (Fig. 280). The ionosphere should not be regarded as a smooth layer causing constant reflection; it is more nearly a turbulent ionic cloudlike region into which the radio waves penetrate a distance and then emerge reduced by absorption. Transmission conditions are ever changing. The situation is well summarized by Dellinger (see footnote, page 517). He states: "The more one views the complexities of radio transmission via the ionosphere, the more he marvels that it provides any intelligible communication."

It is true that *at times* communication can be carried on via the sky-wave path over great distances, and with small amounts of power. Such phenomenal achievements are of much interest, but communication *at times* has little commercial value. Communication at *all* times via the sky-wave path is an impossibility. On the other hand, communication using sky waves can be made quite reliable if the systems have sufficient power, have well-designed

sending and receiving antennas, and use at different times the best frequency for conditions at each daily and seasonal period.

Studies of transmission via the sky-wave path are being made at all times, and much information on the subject is available. From these studies, and from tests made to determine the best locations for transmitting and receiving stations, it is possible to use sky-wave transmission for commercial systems.

The ionosphere is subject to severe "storms," which may last for several days. During such storms the ionosphere appears to be very turbulent, particularly near the earth's magnetic poles. At times the normal layers seem to be broken up, and unstable clouds of ions appear to exist. Transmission by the sky wave is very erratic then, and service may be impossible over a sky-wave path. Such storms appear to be more frequent and pronounced during periods of intense sunspots. Sometimes it is necessary to use a new transmission route. For instance, since the great circle path by which radio waves travel from the east coast of the United States to England passes close to the north polar region, it is sometimes necessary to communicate with England by sending to South America and relaying to England. Sometimes it is necessary to drop to a very low frequency, below the conventional broadcast band.

Interference.—The reception of signals by either the ground wave or the sky wave may be interfered with by undesired signals from other sending stations or by strays. **Strays** are electromagnetic disturbances other than those produced by radio-transmitting systems. These strays may be of two types, (*a*) natural atmospheric disturbances or (*b*) man-made electric disturbances. **Atmospherics** are strays produced by atmospheric conditions, and are usually called **static** in the United States. Distant lightning strokes cause severe static. Man-made electric disturbances are caused by electric lines and equipment. In general, any electric device that sparks, arcs, or causes similar discharges is a source of troublesome radio noise.

Fading.—The strength of the radio signal received undergoes variations in intensity as a result of changes in the transmission path. This is called **fading.** Considering the ground wave only, for the moment, the surface-wave component may experience slight changes in transmission because of variation in the moisture

content of the air, etc. Ordinarily this does not cause fading. Fading often occurs if a sky wave is used. Sometimes reception is by a combined sky wave and ground wave, as shown in Fig. 278. The received signal will be the vector sum of these two waves, and since the sky-wave path is in general erratic, fading may occur. Fading often is largely an amplitude change, and in such instances the automatic volume-control circuit (page 432) will compensate for this. Sometimes, certain bands of frequencies in a radio signal fade, at least partially. This is called **selective fading** and causes distortion.

Antenna Input Impedance.—As was explained in Chap. V, the input impedance of an open-circuited transmission line goes through cyclic variations as the frequency of the impressed voltage is varied. This phenomenon is caused by the interaction of the initial and reflected waves. It is a phenomenon associated with the standing waves on the line. Early in this chapter it was shown that a radio antenna could be developed from an open-circuited transmission line and that a radio antenna had standing waves on it.

It would, therefore, be expected that the input impedance of a radio antenna would vary with applied voltages of different frequencies, and this is true. Unlike an open-circuited line, however, the antenna radiates much power. For this reason the antenna must take power from the transmitter, and hence the standing waves and the variations in input impedance are less pronounced than on an open-circuited line. The input impedance of a broadcast antenna is shown in Fig. 284. As is seen, the antenna input resistance and reactance vary between wide limits. Also, the reactance is inductive at some frequencies and capacitive at others, as would be expected from transmission-line theory.

The antenna of Fig. 284 is resonant at the frequencies of 0.685, 1.070, and 1.715 megacycles, because the reactance is zero at these frequencies. At 0.685 megacycles, the resistance is about 35 ohms, and the load impedance offered by the antenna would accordingly be $Z_L = 35 + j0$. This is a low impedance point and the antenna corresponds at this frequency to a series tuned circuit. This antenna is 350 feet high (see Fig. 284), and is driven at the base. An extensive **ground system** of buried radial wires is used. The current and voltage distribution along the antenna is as indicated in Fig. 285*a*. At 0.685 megacycle, the antenna functions as a

quarter-wave grounded antenna driven at its base. The distance the lines are out from the antenna indicates the magnitude of the current and voltage at each point along the antenna. Because the antenna is radiating power, it must draw power, and the standing waves are not so pronounced as on a lossless open line; that is, the waves do not fall to zero.

At 1.070 megacycles, the resistance is 310 ohms, and the reactance is zero, so that the load impedance offered by the antenna

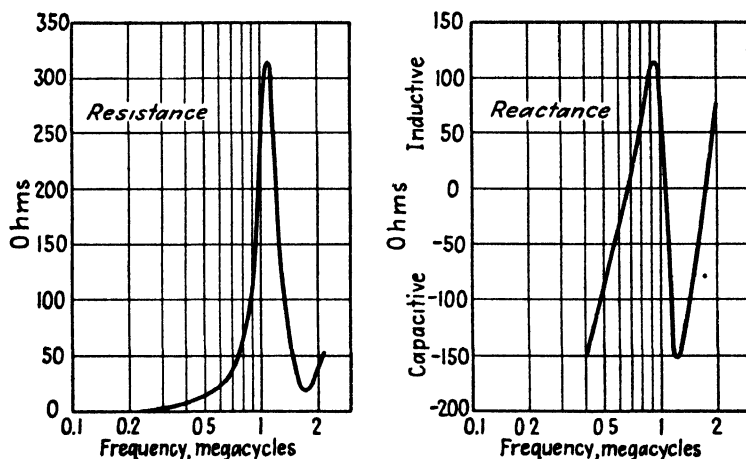



FIG. 284.—Variations of input impedance with frequency for a broadcast antenna. This is for a tower 350 feet high and insulated from ground. The over-all height is 364 feet. The antenna ground system consists of buried conductors extending radially from the base. The antenna input impedance measurements were made between the tower and the ground system These curves are for one tower of a two-tower directional system of Radio Station WEEI, Boston, Mass. This figure is from *The General Radio Experimenter*, Vol. 12, No. 9, February, 1938. (*General Radio Co., and Radio Station WEEI*)

is $Z_L = 310 + j0$. This is a high impedance point, and the antenna corresponds at this frequency to a parallel tuned circuit. The voltage and current distribution is as indicated in Fig. 285*b*. At this frequency the antenna is being operated as an end-driven grounded half-wave antenna. At 1.715 megacycles, the impedance offered by the antenna is $Z_L = 20 + j0$, and the voltage and current distribution are as indicated in Fig. 285*c*. Note that at 1.715 megacycles the antenna again has fallen to a very low impedance, such as that offered by a series resonant circuit.

As will be discussed later, the antenna of Fig. 285 is of the type used in amplitude-modulation radio broadcast over the 550- to

1700-kilocycle range. It is of much interest from the standpoint of radio at higher frequencies to study the input impedance of an antenna isolated from the ground and driven at the center. Such an antenna is shown in Fig. 286.

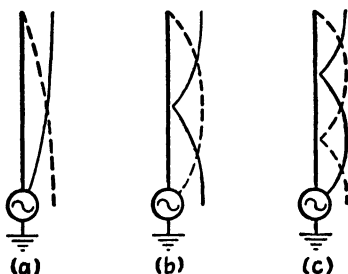



FIG. 285.—Voltage (solid lines) and current (broken lines) distribution along an antenna when operated (*a*) as a vertical quarter-wave grounded antenna; (*b*) as a vertical half-wave grounded antenna; and (*c*) as a three-quarter-wave grounded antenna. The curves indicate the effective values of voltage (solid lines) and current (dotted lines). The values of voltage and current do not drop to zero along the line because the antenna is drawing and radiating power. (See Fig. 258.)

In Fig. 286*a*, the antenna is driven at a frequency such that it is one-half wavelength long, and by analogy with the transmission line (Fig. 258*b*) the voltage and current distribution will be as indicated. *The input impedance will be low*, because at the generator a low value of voltage is forcing a large current into the antenna. In Fig. 286*b* is shown the same antenna driven at a higher frequency such that it is one wavelength long. At this frequency the input impedance will be high, because a large voltage is required to force a small current into the antenna.

The center-driven, or center-fed, half-wave antenna (or dipole) is a widely used device. The theoretical input impedance of an isolated half-wave center-driven antenna is 73 ohms resistance. For such antennas near the surface of the earth, ground reflection causes the input impedance to vary a few ohms above and below this value. If a half-wave antenna is not fed at the center, but is driven at one end, the voltage distribution is as shown in Fig. 285*b*, and the input impedance will be much higher than the 73 ohms which was discussed earlier in this section.

FIG. 286.—When an antenna is driven as indicated in (*a*) at a frequency such that it is one-half wavelength long, the input impedance will be low. When the same antenna is driven at a frequency such that it is one wavelength long as in (*b*), the input impedance will be high.

Because the wave velocity on an antenna is a little less than in free space, and for other reasons, the physical length that a half-wave antenna should have is about 5 per cent less than the theoretical length. Thus at 30 megacycles

the actual length of a half-wave antenna should be about $3 \times 10^8/(2 \times 3 \times 10^7) = 5 \times 0.95 = 4.75$ meters. Above this frequency the factor 6 per cent is used. At a frequency of a few million cycles the factor 3 per cent often is used.

As has been explained, the purpose of a transmitting antenna is to radiate radio-frequency power into space so that communication can be maintained with some distant point. The **radiation resistance** of a transmitting antenna equals the power radiated by the antenna divided by the square of the effective value of the antenna current measured at the point at which the power is applied. The transmitting antenna is not a perfect device, and the power put into the antenna must exceed the power radiated, because of the losses in the antenna and surrounding objects. The **radiation efficiency** of a transmitting antenna is the ratio of the power radiated to the total power supplied at a given frequency.

The losses that occur in an antenna and surrounding objects include (*a*) losses in the surface of the earth in the immediate vicinity of the antenna; (*b*) losses in the conductors that form the antenna (in the most simple case this is the loss in a wire, but in some cases it will be the losses in a complicated tower structure); (*c*) losses in guys; (*d*) losses caused by high-voltage electric discharges known as **corona**; (*e*) losses caused by hysteresis and eddy currents in metallic objects; (*f*) dielectric hysteresis losses in insulators, insulation, wood supports; etc. These losses should be kept at a minimum. In particular, ground losses are reduced by having a buried network of conductors immediately under a grounded vertical antenna and extending radially in all directions for a distance of several hundred feet. This is particularly important with the quarter-wave antenna of Fig. 285*a*, where large currents flow near the antenna base and, hence, in the ground where the "image antenna" is located. The **antenna resistance** is equal to the total power supplied to the antenna divided by the square of the effective value of the antenna current at the point where the antenna is driven. The value of the antenna resistance includes the effect of all the losses enumerated and other losses that may occur.

From the foregoing considerations it is apparent that an antenna is far more than a wire or a tower. In reality it is an electrical network somewhat as shown in Fig. 287. Such a diagram helps to explain why the current at various points along an antenna or

the voltage from antenna to ground at various points may have the distributions shown in Fig. 285.

Broadcast Antennas for Amplitude-modulation Systems.—In radio broadcast, the objective is to provide service for a large number of potential listeners. These may be quite evenly distributed in the various directions from the transmitting antenna. If this is true, and if no other limitations are involved, then the antenna should have a circular horizontal radiation pattern. Service to these potential listeners should be reliable and not subject

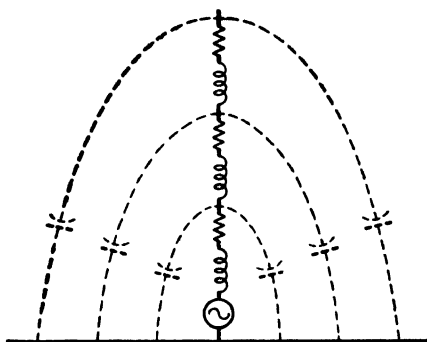



FIG. 287.—An antenna is in reality an electrical network as indicated.

to the erratic variations characteristic of sky-wave transmission. Program sponsors are interested in service to the immediate market area, and are not interested in service to remote localities unless that are advertising on a nation-wide basis.

All these factors require the use of the ground wave for broadcast service. Then the radiation is directed along the earth and not up into the ionosphere. The ground wave must be vertically polarized so that the electric lines of force will be vertical and not horizontal. If this is not done, the energy absorption in the surface of the earth will be great, and the electric field will be attenuated rapidly. As an extreme example of this, the electric field at the surface of a perfect conductor must be zero, because no voltage can exist along a perfect conductor, and a voltage difference is necessary for an electric field to exist.

Nondirectional Broadcast Antennas.—A uniform horizontal radiation pattern with vertical polarization and a strong ground wave is obtained by an antenna like Fig. 285 *a* or *b*. Hence, such antennas are used extensively in ordinary radio broadcast. As

shown in Fig. 276, the radiation pattern in a *horizontal plane* along the surface of the earth for a single vertical antenna is a circle, and such an antenna is said to be a **nondirectional antenna.**

The radiation pattern in the *vertical plane* for such an antenna is shown in Fig. 274, if the antenna is one-half wavelength long and the lower end is at the surface of the earth. With this antenna the sky-wave radiation is small compared with the radiation along the surface of the earth. Such an antenna ordinarily is supported on large insulators, and is driven by applying the voltage from the transmitter between the antenna and ground system. As mentioned elsewhere in this chapter, a system of radial ground wires is used to reduce the energy loss at the base of the antenna. Because of the fact that vertical antenna is not a simple straight wire but is a fabricated tower, and for other reasons, the actual antenna height must be about 0.625λ instead of 0.5λ for best results. The symbol λ represents the wavelength of the signal to be radiated.

Directional Broadcast Antennas.—It sometimes happens that the population distribution is not uniform with respect to an antenna, and that the signal energy should not be radiated uniformly, but should be radiated in certain directions. Or it may be desirable to decrease the strength of the signal radiated in a given direction so as to reduce the interference that might be caused by some other station operating on the same frequency. Radiation in a given direction, or directions, is accomplished with **directional antennas.**

Directional horizontal radiation patterns for broadcast purposes are produced by using two or more vertical antennas. The pattern is influenced by the spacing, the relative magnitudes of the currents fed the two towers, and the phase angles between these two currents. Innumerable combinations are possible, and the shapes of the patterns are many and varied.

Directional patterns are possible because of the way that the signals from the separate antennas combine in various directions from the antennas. Thus suppose that two antennas *A* and *B* are one-half wavelength apart, that the currents are equal in magnitude, and that the antennas are driven in phase. That is, when the current in one antenna is a maximum value, the current in the other is a maximum value. As indicated in Fig. 288*a*, each antenna is sending out waves, and for convenience the peak values of the positive crests of the waves are shown by solid circles, and the peaks of the negative half cycles are shown as broken lines. At points

along the line x-y passing through the antennas, cancellation occurs because the positive and negative values from the two antennas arrive at the same point at the same time and neutralize each other. At all points along the line M-N at right angles to the antennas the signals arrive at a given point so that they add. Thus the radiation pattern will be as shown in Fig. 288*b*.

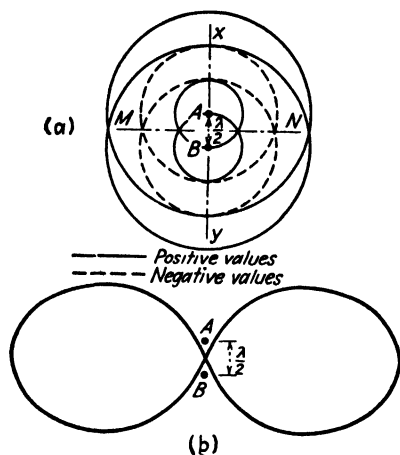



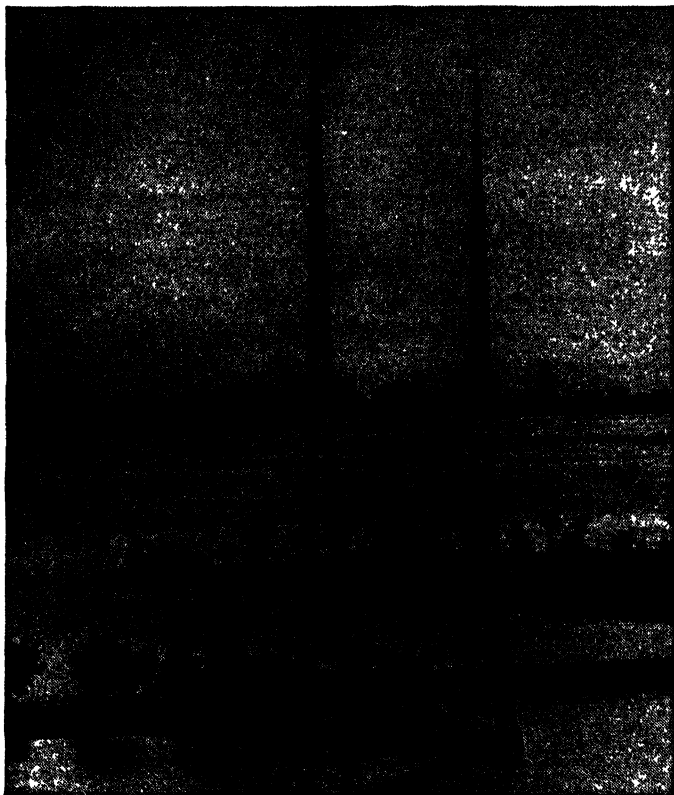
FIG. 288.—In (*a*) is shown how the waves from two vertical antennas one-half wavelength apart and driven in phase by equal currents act to cause cancellation of radiation in directions along the line of the antennas, but add to produce strong radiation at right angles to the antennas. The calculated pattern is shown in (*b*).

Design of a Directional Broadcast Antenna.—The method used in determining the radiation pattern in Fig. 288 is not sufficiently general for many cases, a vector solution being required. This vector solution will be explained now, using the antenna of Fig. 288 as an illustration.

The antenna consists of two vertical radiators spaced one-half wavelength apart ($\lambda/2 = 180°$) and with currents having identical magnitudes and phase relations in each radiator. The antenna radiation pattern is found by determining the relative strengths of the total combined signal from the two radiators A and B at some distant point P, as in Fig. 289. This point is so remote that the two rays are assumed parallel. The radiation from antenna B must travel a greater distance than the radiation from antenna A to reach distant point P. For this reason the two signals will arrive out of phase by an amount $S \sin \theta$, where S is the spacing in degrees and θ is the angle made with a line at right angles to the radiating towers A and B.

FIG. 289.—To reach distant point P the radiation from antenna B must travel a distance $S \sin \theta$ farther than the radiation from antenna A.

When $\theta = 0°$, $S \sin \theta = 0$, and the waves from the two antennas will arrive in phase at a point located on a line at right angles to

the radiators; hence, the two signals add numerically. If it is arbitrarily stated that the radiation for each tower alone is 1.0, then the combined signal strength in the direction $\theta = 0$ is 2.0. At an angle $\theta = 30°$, $S \sin \theta = 180° \times 0.5 = 90°$, and the signal from

The two towers provide a directional radiation pattern for an amplitude-modulation radio-broadcast station. The towers are 300 feet high. Salt water provides an excellent ground. (*Radio Station WIOD, Miami, Florida, and the Blaw-Knox Co.*)

radiator B will arrive at a distant point along the $\theta = 30°$ line 90° behind the signal from radiator A. The combined signal strength at this angle will be the vector sum of two unit radiations 90° apart, or $\sqrt{1.0^2 + 1.0^2} = 1.414$. At an angle $\theta = 60°$, $S \sin \theta = 180 \times 0.866 = 156°$, and the signal from radiator B will arrive at distant point along the $\theta = 60°$ line 156° behind the signal

from radiator A. The combined signal strength at this angle will be the vector sum of two unit radiations 156° apart, or

$$\sqrt{(1.0 + 1.0 \cos 156°)^2 + (1.0 \sin 156°)^2} =$$
$$\sqrt{[1.0 + (1.0 \times -0.91)]^2 + (1.0 \times 0.41)^2} =$$
$$\sqrt{(1.0 - 0.91)^2 + (0.41)^2} = \sqrt{(0.09)^2 + (0.41)^2} = 0.42 \text{ units.}$$

At an angle $\theta = 90°$, the phase relation will be 180°, the two received components will cancel, and the radiation in the direction of the two radiators will be zero, as Fig. 288 shows. This pattern shows the radiation in various directions along the surface of the earth.

It is doubtful if a pattern such as Fig. 288 will ever be desired. It is more probable that a pattern is to be "bean-shaped" so that the local listeners will be served, and also so that listeners located up and down a valley will receive strong signals. Such patterns can be produced because, as previously mentioned, the magnitudes of the currents in the two towers and their phase angles can be altered. The pattern of Fig. 288 was for identical currents and zero phase angle. The design of an antenna for a more practical "bean-shaped" pattern will be considered.

Illustrative Problem.—It is desired to design a two-element antenna system which will be located at the edge of a city, and which will serve the radio listeners close to the antenna and also direct much signal energy up and down a densely populated valley.

Solution.—Step 1. Decide on an ideal pattern by examining the geographical population distribution and other pertinent factors. Then consult a chart showing various patterns[1] and select the combination giving the most suitable pattern. Thus the chart consulted shows that an antenna system composed of two vertical radiators 1 5° apart, with currents of equal magnitude but phased 45°, produces a pattern almost suitable. From here on the design largely is a cut-and-try process. A few calculations (made as will be explained) indicate that the desired pattern will be obtained if the spacing is 135°, the phasing such that the current in radiator B lags that in A by 50°, and the current ratio is such that if the current in A is 1.0, that in B is 0.7.

Step 2. The radiation pattern now will be calculated by determining the radiation component from each antenna, and then combining these components vectorially to find the total radiation.

[1] Such charts are given by R. M. Foster, Directive Diagrams of Antenna Arrays, *Bell System Technical Journal*, Vol. 5, April, 1926; also, by G. H. Brown, Directional Antennas, *Proceedings of the Institute of Radio Engineers*, Vol. 25, No. 1, January, 1937. Such information is summarized by F. E. Terman, "Radio Engineers' Handbook," McGraw-Hill Book Company, Inc.

When $\theta = 0°$, $S \sin \theta = 135° \times 0 = 0$. The two waves travel the same distance but do not arrive in phase because of the fact that the current in radiator B lags that in A by 50°; also, the current ratio is $B = 0.7A$. Hence, at right angles to the antennas the field strength from radiator A will be 1.0, and that from B will be 0.7 and will lag by 50°. The combined signal strength will be the vector sum, or

$$\sqrt{[1.0 + (0.7 \cos 50°)]^2 + (0.7 \sin 50°)^2} = \sqrt{[1.0 + (0.7 \times 0.64)]^2 + (0.7 \times 0.77)^2} = 1.55 \text{ units.}$$

When $\theta = 30°$, $S \sin \theta = 135° \times 0.5 = 67.5°$. The wave from radiator B must travel 67.5° farther, and since it was radiated 50° lagging, the total angle of lag will be $67.5° + 50° = 117.5°$. Hence, at the 30° angle indicated on Fig. 290, the combined signal strength will be

$$\sqrt{[1.0 + (0.7 \cos 117.5°)]^2 + (0.7 \sin 117.5)^2} = \sqrt{[1.0 + (0.7 \times -0.46)]^2 + (0.7 \times 0.89)^2} = 0.92 \text{ units.}$$

When $\theta = 60°$, $S \sin \theta = 135° \times 0.866 = 117°$. Adding 50° gives 167° that the radiation from B lags that from A. The combined strength is

$$\sqrt{[1.0 + (0.7 \cos 167°)]^2 + (0.7 \sin 167°)^2} = \sqrt{[1.0 + (0.7 \times -0.97)]^2 + (0.7 \times 0.22)^2} = 0.36 \text{ units.}$$

When $\theta = 90°$, $S \sin \theta = 135° \times 1.0 = 135°$. Adding 50° gives 185° that the radiation from B lags that from A. The combined strength is

$$\sqrt{[1.0 + (0.7 \cos 185°)]^2 + (0.7 \sin 185°)^2} = \sqrt{[1.0 + (0.7 \times -0.99)]^2 + (0.7 \times 0.09)^2} = 0.31 \text{ units.}$$

Step 3. The preceding calculations have given the shape of the radiation pattern in the first quadrant from 0 to 90°. The shape of the radiation pattern in the fourth quadrant from 270° to 360° now will be found. It is not necessary to make the calculations for the second and third quadrants because the pattern will be symmetrical.

When $\theta = 270°$, a diagram similar to Fig. 289 but with $\theta = 270°$ will show that the radiation from A must travel 135° farther, but since the current in B is 50° behind that in A, the signal from A will arrive at a point along the 270° line $135° - 50° = 85°$ later than that from B. The radiation along the 270° line is then composed of two vectors, one of 1.0 units long, and one 0.7 units long leading the first by 85°. The resultant will be

$$\sqrt{[1.0 + (0.7 \times \cos 85°)]^2 + (0.7 \sin 85°)^2} = \sqrt{[1.0 + (0.7 \times 0.09)]^2 + (0.7 \times 0.99)^2} = 1.27 \text{ units.}$$

When $\theta = 300°$, $S \sin \theta = 135° \times 0.866 = 117°$. The radiation from A must travel 117° farther, but subtracting the 50° phase angle gives the radiation from A arriving at a distant point on the 300° line an angle of 67° later than the radiation from B. The radiation is thus the sum of one vector 1.0 unit long and a vector 0.7 unit long but leading by 67°, that is,

$$\sqrt{[1.0 + (0.7 \times \cos 67°)]^2 + (0.7 \sin 67°)^2} = \sqrt{[1.0 + (0.7 \times 0.39)]^2 + (0.7 \times 0.92)^2} = 1.43 \text{ units.}$$

When $\theta = 330°$, $S \sin \theta = 135° \times 0.5 = 67.5°$. Subtracting 50° gives 17.5° as the angle by which the 0.7 unit radiation from *B* leads the 1.0 unit radiation from *A*. Adding the vectors gives

$$\sqrt{[1.0 + (0.7 \times \cos 17.5°)]^2 + (0.7 \sin 17.5°)^2} = \sqrt{[1.0 + (0.7 \times 0.95)]^2 + (0.7 \times 0.30)^2} = 1.67 \text{ units.}$$

The final radiation pattern is plotted in Fig. 290.

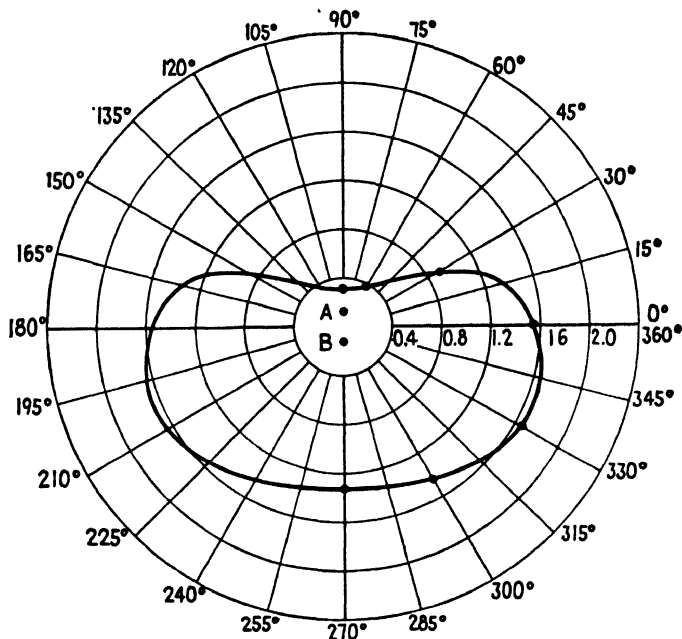



Fig. 290.—Radiation pattern in a horizontal plane at the surface of the earth for two vertical antennas *A* and *B* of the conventional broadcast type. Antenna *B* is 135° from antenna *A*. The current in antenna *B* is 0.7 that of antenna *A*, and lags by 50°. The dots on the curve are values calculated as explained in the text.

A system of antennas coupled together to obtain directional effects is called an **antenna array**. Antenna arrays of two or more radiators are common in broadcast. The same general method just explained can be used on arrays of more than two radiators. Both mechanical and electrical radiation-pattern calculators have been devised.[1]

This section has considered horizontal radiation only. Radia-

[1] As an example of a mechanical calculator, and for additional information on pattern calculations, consult F. A. Everest and W. S. Pritchett, "Horizontal-polar-pattern Tracer for Directional Broadcast Antennas," *Proceedings of the Institute of Radio Engineers*, Vol. 30, No. 5, May, 1942.

tion in the various vertical planes can be computed by determining the signals that arrive from the several antennas and by combining these vectorially. Thus it is possible to construct a three-dimensional radiation pattern, such as Fig. 291, which represents

FIG. 291.—Model showing radiation in the horizontal plane, and in various vertical planes for the antenna described in Fig. 290. (*Radio Station KOAC.*)

radiation in all possible directions from an antenna similar to the one considered in the preceding problem.

Antennas for Point-to-point Communication.—Directional transmitting antennas are used for this service. For transoceanic purposes, the high-frequency short-wave band from 3 to 30 megacycles, or 100 to 10 meters, is used. Two types of transmitting

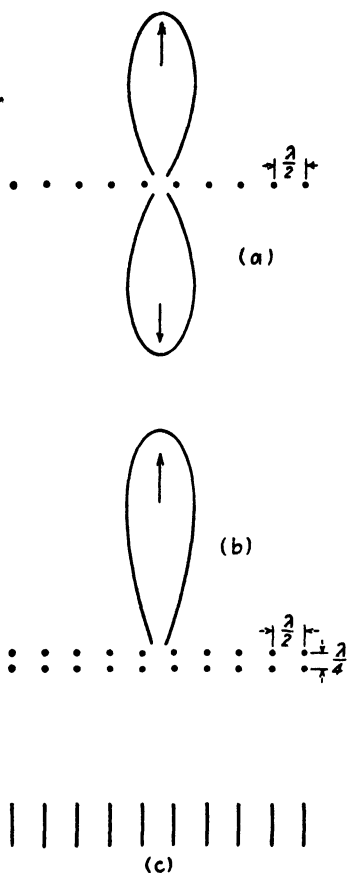



FIG. 292.—In (*a*) are shown the tops of a row of vertical radiators. The radiation in a horizontal plane will be in two directions as indicated if the radiators are spaced one-half wavelength apart and are driven in phase. In (*b*) is shown a second row of parasitically excited radiators placed one-quarter wavelength behind the first row. The radiation now largely occurs in one direction. The radiation patterns are *not* to scale. In (*c*) is shown a front view of the tier array of radiators.

antennas have been employed: (*a*) antenna arrays consisting of "stacks" or tiers of relatively short antennas, each of which is perhaps a half-wavelength long, and (*b*) long-wire antennas composed of several conductors, each of which is perhaps several wavelengths long.

Tier Arrays.—The two vertical radiators of Fig. 288, which were one-half wavelength apart and driven in phase, were found to have directional radiating properties. If more radiators are added, an extension of the method explained in Fig. 288*a* will show that the radiation is still zero in the direction in which the row extends, and that the radiation is strengthened at right angles or "broadside" to the row of vertical radiators. For a row of, say 10 radiators very strong radiation will occur in both directions, as Fig. 292*a* indicates.

For point-to-point communication, back radiation would be undesired, because the signal energy sent out the back would be wasted, and might cause interference. To direct all the energy toward the distant receiving station, a second row of radiators is placed directly behind the first row. The distance between the two rows is one-fourth wavelength, or 90°. The second or back row is not connected to the generator, but is excited by induction, or "parasitically" as it is called.

The action of the parasitically driven back row of Fig. 292*b* can be explained as follows: Assume that a given conductor in the front

row radiates a positive impulse. This will travel one-fourth wavelength, or 90°, to the back row, where by induction a current will be set up in a back radiator. This current will not be in a direction to *assist* the impulse that produces it, but will be in the direction to oppose the impulse producing it. This means that the back radiator will send out an impulse 180° out of phase with the original *positive* impulse that excited it. This radiated *negative* impulse will travel toward the front conductor that is one-fourth wavelength, or 90°, away. During the 90° interval taken for the original impulse to travel from the front radiator to the back radiator, and the 90° interval required for the parasitically excited impulse to travel from the back radiator to the front radiator, the voltage impressed on the front radiator has passed through 180°, and the front radiator is transmitting a negative impulse when the negative impulse from the back radiator arrives. These signals add, and they strengthen the radiation out the front side of the antenna. On the back side of the antenna system the radiation from the parasitic radiators is 180° out of phase with that from the front radiators, and cancellation occurs. Thus the back row acts like a reflecting mirror and reflects the energy back out the front side.

The diagrams of Figs. 292*a* and *b* are as viewed from the top, and show the radiation in a horizontal plane. An observer standing in front of the system would see the vertical conductors as shown in Fig. 292*c*. Such a system would have a vertical radiation pattern somewhat as shown by Fig. 293*c*, and would therefore direct sky-wave energy up toward the ionosphere, particularly at low angles. Several tiers of vertical radiators one above the other have been used.

Antenna systems, such as the ones shown in Fig. 292, have had extensive use, but are intricate, are expensive, and operate at one frequency only. This is of importance in transoceanic systems, where each channel is provided with several different frequencies on which to operate as ionospheric conditions change.

Rhombic Antennas.—These are composed of four conductors, each several wavelengths long, arranged as a rhombus, as shown in Fig. 293*a*. This antenna is supported above the earth by any convenient means, for instance, by "telephone" poles. The plane of the antenna is parallel to the earth.

The conductors of Fig. 293*a* are *not* resonant, and there are no pronounced standing waves along them. In fact, the wires are

terminated in a resistance R that largely prevents standing waves. But the current decreases logarithmically along the wires, because they are, in a sense, transmission lines. The radiation pattern from each wire is somewhat as indicated, minor radiation lobes having been omitted.

It will be noted that one of the major lobes of each wire points in the same direction, and, hence, radiation combines in the direc-

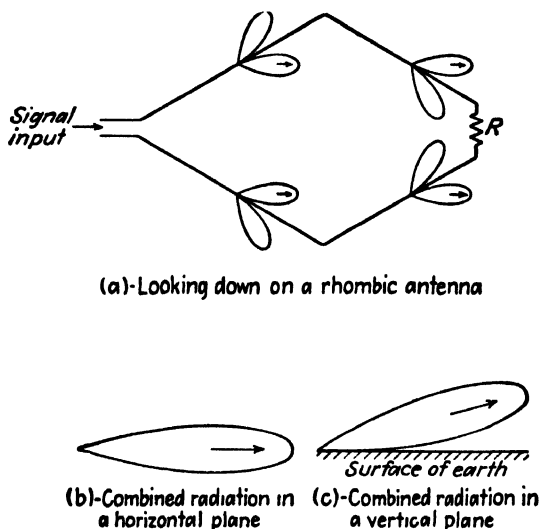
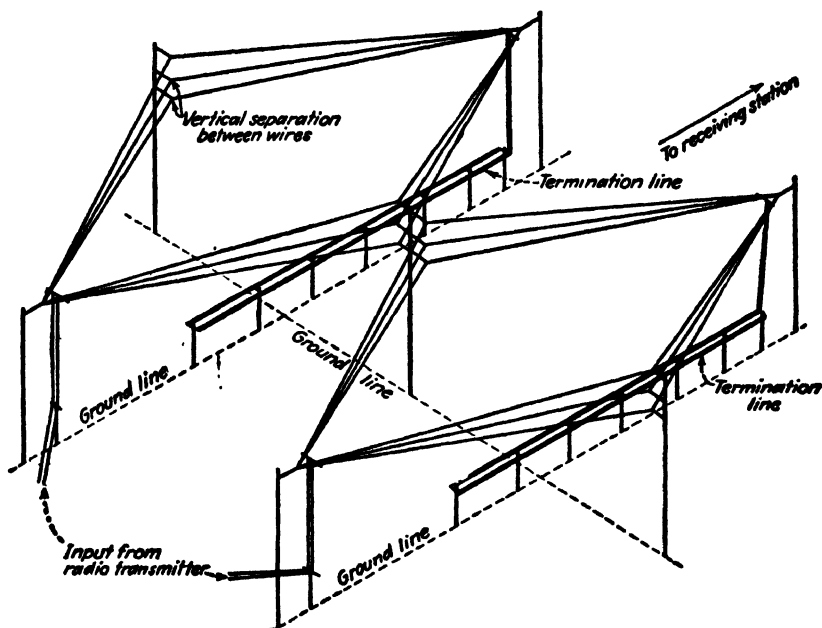



Fig. 293.—The rhombic antenna (*a*) is composed of four wires each several wavelengths long. Each wire has a radiation pattern of the form shown in (*a*), the minor lobes having been omitted. The wires are so arranged that the radiations from each wire in the direction of the arrows all add, giving a strong horizontal radiation pattern as shown in (*b*). Radiations in the other directions (not shown by arrows) largely cancel. The vertical plane radiation is shown by (*c*). Resistance R is a termination much as on a transmission line to prevent wave reflection. Although much energy is dissipated in R, the rhombic antenna is one of the most practicable of all directional antennas. These figures are *not* to scale.

tion of the arrow. The directions of the other lobes are such that radiation cancels. As a result, the directional pattern is as shown in Fig. 293*b*. In the vertical plane the radiation is upward at an angle that can be controlled by design factors.[1] This is of particular importance in short-wave communication over great distances by sky-wave radiation. This antenna is inexpensive, will work over a wide frequency range, and is used extensively.

[1] A good discussion will be found in "The A.R.R.L. Antenna Book" by the American Radio Relay League.

Metallic Reflectors.—For very high frequencies (30 to 300 megacycles), and for point-to-point service, it is possible to use antennas similar to those just considered. In particular, driven half-wave antenna rods with parasitic reflectors are popular.

For ultrahigh frequencies (300 to 3000 megacycles), and above, the reflectors usually take the form of sheet-metal reflectors shaped

A twin rhombic transmitting antenna used in transoceanic radio-telephone circuits for talking over distances of several thousands of miles. This antenna produces highly directional transmission in the general range of frequencies of from 5 to 25 megacycles. The termination line often is of iron wire and serves as resistor R of Fig. 293*a*. (*American Telephone and Telegraph Co.*)

as shown in Fig. 294. The antennas often are center-driven half-wave radiators located at the focal point. Of course at ultrahigh frequencies such half-wave radiators are very short. Sometimes wire-mesh reflectors instead of sheet-metal reflectors are used to decrease wind resistance, and for other practical reasons. Of course the meshes must be close together for good results.

The focusing action of such metallic reflectors is similar to that of light reflectors. The wave energy radiated backward cannot penetrate the metal sheet, or screen, and is redirected outward.

If the shape of the reflector is correct, a beam radiation of closely parallel rays can be obtained.

Horn Radiators.—Such radiators are used to some extent at ultrahigh frequencies, and particularly at superhigh frequencies (3000 to 30,000 megacycles). These are usually straight-sided horns of rectangular cross-sectional area, and very sharp beams of

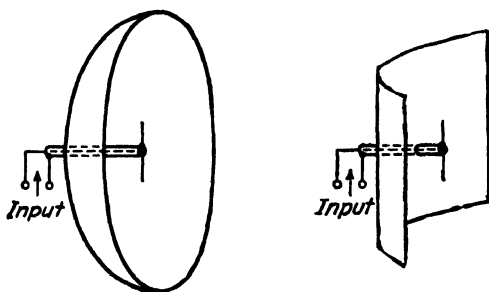



FIG. 294.—Two convenient forms of reflectors used at ultrahigh frequencies. The signal energy is fed in through a coaxial cable. The center conductor is bent down forming one half of a half-wave antenna. A short piece of wire attached to the outer cable conductor forms the other half of the antenna. Such reflectors often are made of a wire mesh.

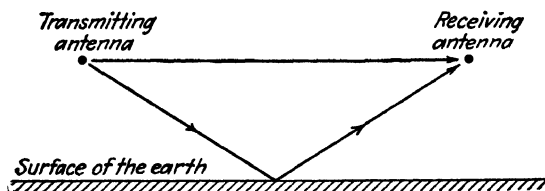



FIG. 295.—Over a line-of-sight path the signal arrives as a direct ray and as a reflected ray.

electromagnetic energy can be radiated with them. These horns are fed by **wave guides,** which are hollow metallic tubes which guide the electromagnetic waves within from the source of energy to the horn. The energy at these extremely high frequencies often is generated by magnetrons (page 394). A coaxial-cable-fed half-wave antenna located inside a wave guide readily transfers energy in the form of electromagnetic waves to the guide.

Radio Communication by Line-of-sight Path.—It has been shown that there is a maximum usable frequency for point-to-point communication. Thus frequencies above *about* 30 megacycles (depending on conditions) are not satisfactory for sky-wave communication, and because of excessive losses are not suitable for

ground-wave communication. They can, however, be used for point-to-point communication over a "line-of-sight" path, as shown in Fig. 295. This neglects the curvature of the earth.

As is indicated, the signal at the receiver arrives by two paths. At these very high frequencies there is a 180° phase shift at the point of reflection for both vertically and horizontally polarized waves. (At low frequencies only horizontally polarized waves are shifted 180°.) If the paths of Fig. 295 are of the same length, as they are when the antennas are close to the ground, the two waves will arrive 180° out of phase and will cancel. If the antenna heights and the distances apart are such that the paths differ by one-half wavelength (which is not much in terms of feet at these frequencies), then the two-signal rays arrive in phase, and add to produce a strong signal. The importance of a little experimentation in locating transmitter and receiver sites is evident. Of course, if the signal at the transmitter could be formed into a beam of parallel "searchlight" rays, then the reflected ray could, theoretically, be ignored.

An eight-element "square-loop" antenna for frequency-modulation broadcasting in the 100-megacycle region over a line-of-sight path. This antenna radiates a horizontally polarized signal in all directions in a horizontal plane and suppresses radiation at vertical angles. (*Federal Telephone and Radio Corporation, manufacturing subsidiary of International Telephone and Telegraph Corporation.*)

Either vertical or horizontal antennas can be used for communication by line-of-sight path. A simple straight-wire vertical antenna will radiate a **vertically polarized field,** because the voltage difference is along the vertical antenna, and the resulting electric lines of force will be parallel to the antenna and, hence, vertical. The direction of the electric lines of force is by definition the **direction of polarization.** It follows that a simple straight-wire horizontal antenna radiates a **horizontally polarized field.** A

vertical field will induce a signal voltage only in a vertical antenna, and a horizontally polarized field will induce voltage only in a horizontal antenna. It is important to note, however, that a shift in polarization often occurs between the transmitter and the receiver. Vertical antennas often are used for line-of-sight communication.

The importance of height above the ground for the transmitting and receiving antennas also is shown by Fig. 296. The distance D is given by the equation[1]

$$D = 1.41\sqrt{x},$$

where D is the distance in miles over level terrain and x is the height in feet of the transmitting antenna. This equation includes

FIG. 296.—Showing how the curvature of the earth interferes with the transmission of signals over a line-of-sight path.

the effect of refraction in the lower atmosphere. This bends the waves somewhat and increases the distance over which communication is possible, making the distance greater than the actual line-of-sight path.

Antenna Feeders.—In some instances a transmitting antenna is connected directly to the radio transmitter, but often the antenna is located away from the transmitter where the radiation will be unobstructed. In these instances the energy must be transmitted over an antenna-feeder system between the radio-transmitting set and the associated antenna.

The two common circuits for connecting a transmitting set to the transmitting antenna are (*a*) the open-wire transmission line and (*b*) the coaxial cable. The basic types of lines and cables were discussed in Chap. V.

Transmission lines are operated in one of two ways, either as mismatched lines with standing waves on them (also called **resonant, or tuned, lines**) or as matched lines with no standing waves

[1] See footnote, p. 536.

on them (also called **nonresonant lines, untuned lines,** or **flat lines**).

Mismatched, or Resonant, Lines as Feeders.—If the input impedance of a transmitting antenna does not equal the characteristic impedance of the feeder line, then standing waves will exist on this line. Of course these standing waves will not be so pronounced as they would be if the line were open- or short-circuited instead of being connected to the antenna and delivering power to it.

FIG. 297.—Two methods of feeding half-wave antennas *A-B*. In (*a*) the antenna input impedance connected to the feeder or transmission line is low. In (*b*) the line is terminated in a high impedance. The dotted lines are the spreaders, as the line insulators are called. Standing waves will exist on the transmission lines, but these have not been shown.

Standing waves on an antenna-feeder line are not objectionable, and often it is desirable to operate a feeder in this way, as in the following example: Suppose that a radio system is to operate at different times at one of several frequencies using the same transmitting antenna. The antenna input impedance depends on the driving frequency (page 523), and, hence, this input impedance will change when the transmitting frequency is changed. If an impedance-matching network (page 112) were placed between the line and antenna, this network would need to be changed each time the frequency was shifted. But if the feeder transmission line is allowed to operate as a mismatched line with standing waves on it, then no matching network is needed at the antenna, and no adjustments need be made at the antenna when a band is changed. Also, it sometimes is difficult to place a tuning network between the line and antenna, particularly if the antenna input terminals are elevated. Two methods of feeding elevated half-wave antennas with mismatched lines are shown in Fig. 297.

Voltage curves E and current curves I are shown along the antennas. If the half-wave antenna is center-fed, it is driven at a low-voltage high-current point, and the input impedance $Z = E/I$ will be low, about 73 ohms resistance. If the same antenna is end-fed, the input impedance it presents to the feeder will be much higher.

Now the feeder may be mismatched to the antenna, but it must be connected in the proper manner to the transmitter or it will not

load the transmitter correctly. This means that the feeder must be connected to the transmitter through a matching or impedance-transforming circuit (page 112). The type of circuit used for this purpose depends on the input impedance to the mismatched feeder *at the point where it is to be connected to the transmitter.* This fact follows from the theory in Chap. V. For instance, a feeder transmission line, composed of two No. 12 A.W.G. copper conductors, spaced 6 inches apart, will have a radio-frequency characteristic

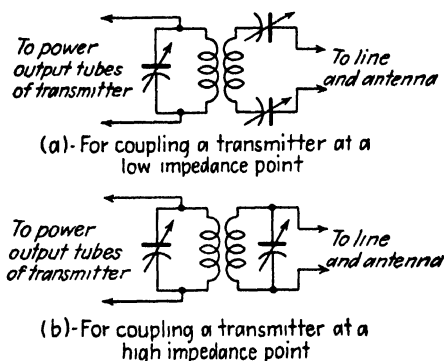



FIG. 298.—When a feeder (transmission line) is connected to an antenna with *no* impedance-matching network, a mismatch exists (usually) and standing waves will be on the feeder. The impedance that is offered to the radio transmitter will depend on the length of the feeder. For connecting to a length offering low impedance, circuit (*a*) should be used. Circuit (*b*) is used for connecting to a length offering high impedance. These circuits match the line to the transmitter so that power is drawn and fed to the antenna.

impedance of about 600 ohms resistance. If this is used to center feed a 73-ohm half-wave antenna, then the line will be terminated in what *approaches* a short circuit. At odd-multiple ($\frac{1}{4}\lambda$, $\frac{3}{4}\lambda$, etc.) quarter-wave points, the input impedance to the mismatched line will be high (above 600 ohms), but at even-multiple ($\frac{1}{2}\lambda$, λ, etc.) quarter-wave points, it will be low (below 600 ohms). Also, if the line and connected antenna are driven at exact quarter-wave line multiples, the input impedance will be pure resistance. For other lengths, it will be resistance and reactance.

If the line is driven by the transmitter at a low-impedance point, then the impedance-transforming circuit of Fig. 298*a* can be used. The two condensers tune the circuit to resonance, and through the mutual coupling the reflected impedance (resistance) is so adjusted that the tubes in the output of the transmitter are loaded properly.

If the line is driven by the transmitter at a high-impedance point, then a circuit such as Fig. 298*b* should be used. This circuit also can be adjusted to couple the proper resistance into the plate circuit of the transmitter output tubes, following coupled-circuit theory (page 102).

Matched, or Nonresonant, Lines as Feeders.—If a radio system is to transmit *at one frequency only*, the feeder often is matched to the antenna so that there are no standing waves on it. The power efficiency of such a matched feeder is somewhat greater, and the radiation is less, than for a mismatched feeder. The so-called "delta-matching transformer" of Fig. 299 sometimes is used for matching at the antenna. The end of the transmission line is flared out and connected so that a match is obtained. This condition can be ascertained with a small neon lamp bulb or a milliammeter, as explained in Chap. V.

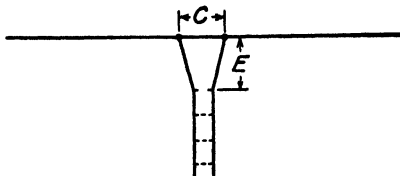



FIG 299—Method of matching a transmission line or feeder to a half-wave antenna With the proper adjustments no standing waves will exist on the line.

This so-called "delta-matching transformer" is in fact a form of tapered transmission line, which in turn is a form of nonsymmetrical network which has impedance-transforming properties. The flared-out section shows different input impedances (*image impedances* are more correct) when measured from either end, and this makes the matching possible. The dimensions for matching are given[1] as

$$E = \frac{123}{f} \quad \text{and} \quad C = \frac{148}{f},$$

where E and C are in feet and f is in megacycles. This is for matching a 600-ohm feeder, composed of No. 12 A.W.G. wires, spaced 6 inches apart, to a half-wave antenna.

Open-wire transmission lines often are matched to antennas by the use of resonant quarter-wave sections (page 159) and **stubs.** These stubs are short sections of the transmission line connected to the line at certain points. The theory is as follows: If a transmission line feeder is mismatched, then standing waves will exist along it, as shown in Fig. 300*a*. The extent to which these waves

[1] See footnote, p. 536.

exist will depend on the degree of mismatch. This is true, because no mismatch causes no standing waves, and a complete mismatch, such as an open or a short circuit, causes complete standing waves that rise to a maximum and fall to zero. The ratio of voltage maximums to voltage minimums, or current maximums to current minimums, on a mismatched line is a measure of the degree of mismatch. The stub is a short section (not a quarter-wave multiple) of an open, or short-circuited transmission line and is therefore a reactance (page 156). The matching is accomplished by connecting this reactive stub across the transmission line at a point such that the stub, and the line and antenna beyond the stub, are in resonance, and also the resistance of the line and antenna beyond the stub are transformed to the proper value to match the transmission line at the point where the stub is connected.

FIG. 300.—Showing how a stub can be used to match an antenna to a line. Between the stub and the radio transmitter the standing waves largely are eliminated.

A short-circuited stub is shown attached to the line of Fig. 300*b*. As indicated, a mismatch and standing waves exist between the antenna and stub, but not between the stub and the transmitter. The location of the stub can be found from Fig. 301. As an illustration, if the ratio of the maximum voltage to the minimum voltage is 3.0 when the line is connected to the antenna *without* the stub, then a short-circuited stub 0.114 wavelength long located 0.170 wavelength from the voltage maximum point toward the transmitter will reduce the standing waves to substantially zero between the stub and the transmitter.

The matching circuit at the transmitter end is quite simple if the line is matched to the antenna with a stub or otherwise. In this instance the input impedance at the transmitter end will be the characteristic impedance of the line, which is resistance at radio frequencies. The circuits of Fig. 302 commonly are used. These are designed in accordance with the theory in Chap. IV.

Coaxial Cables as Feeders.—These cables (page 141) are used ex-

tensively for driving vertical broadcast antennas, one example being given in Fig. 303*a*. The antenna here considered is a base-insulated quarter-wave antenna, and it has an input impedance at the base of about 36 ohms. The coil and condenser selected are of such values that the input impedance is changed (often) to 77 ohms to match the coaxial cable. At the transmitter end the

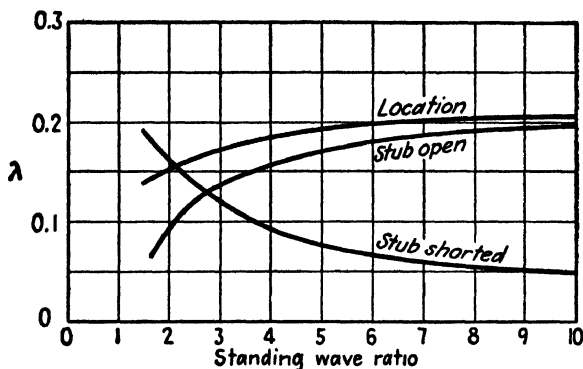



FIG. 301.—Curves for designing matching stubs such as shown in Fig. 300. The curve marked "location" gives the location from a voltage maximum toward the transmitter if a shorted stub is used, and from a voltage maximum toward the antenna if an open stub is used. The curves marked "stub open" and "stub shorted" are the lengths of stubs to be used. All lengths given are fractions of a wavelength. It is assumed that the characteristic impedance of the stub equals that of the line.

cable can be matched to the power-output tube with an air-core transformer following coupled-circuit theory (page 101). A **Faraday shield** often is inserted between the windings of such a transformer. This often takes the form of an arrangement of parallel conductors (wires) with all conductors tied together at one end and grounded. At the other end the wires are not connected. This prevents the grounded unbalanced coaxial cable from capacitively unbalancing the load circuit of the tube. Magnetic lines of force can pass through the shield and induce the required signal voltage in the transformer secondary, but it is so oriented that electric lines cannot.

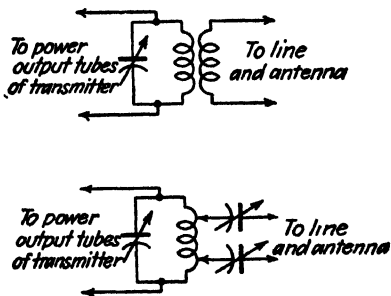



FIG. 302.—Matching networks for connecting matched or nonresonant lines to radio transmitters.

Another method of feeding a vertical broadcast antenna is shown in Fig. 303*b*. In this system the antenna is grounded at the base instead of being supported on insulators. This method will be recognized as being "one-half" of the "delta transformer" previously described. The characteristic impedance of a coaxial cable often is 77 ohms (page 145), and for a match the input impedance at the lower end of the sloping antenna feed wire must be 77 ohms. A high-frequency impedance bridge, driven at the radio frequency to be used, is connected to the lower end of the wire. The upper end of the wire is moved along the antenna until the resistance component measured by the bridge is 77 ohms. Then a condenser at the lower end of the wire is adjusted to neutralize the inductance that usually appears with the sloping wire arrangement used. When the sloping wire with the series condenser is connected to the coaxial cable, an impedance match exists, and no standing waves will be on the cable. The radio-frequency bridge must be used, since the voltage, or current, along the cable cannot conveniently be measured because of the sheath. The coaxial cable can be matched to the transmitter by the general method explained in the preceding paragraph.

FIG. 303.—Methods of connecting coaxial cables to typical antennas. Diagram (*a*) is for a base-insulated antenna, and diagram (*b*) is for a grounded antenna.

SUMMARY

A transmitting antenna is a conductor, or system of conductors, for radiating radio waves into space. The transmitting antenna can be developed from a transmission line. In the line, the current and voltage components cancel along the line and little radiation occurs, but when the ends of the line are opened out into an antenna the components do not cancel, and radiation occurs.

The field established by an antenna consists of two parts, an induction field and a radiation field. The induction field exists only in the immediate vicinity of the antenna but the radiation field extends out into space and transmits the radio-wave energy.

An antenna in free space has a certain radiation pattern, the shape of which depends on the length of the antenna and the plane in which the radiation is considered, or "viewed." When an antenna is near the surface of the earth, or near some other large reflecting surface, the antenna radiation pattern is

changed because of reflection. The actual radiation pattern of an antenna in the vicinity of the earth is determined by modifying the free-space pattern by the proper reflection factors.

Radiation from an antenna may be divided into two parts, sky wave and ground wave. The sky wave is directed by the antenna up toward the ionosphere, where it may be reflected back to the earth. The ground wave is composed of a surface wave that travels along the surface of the earth and a space wave that itself consists of two components. One component of the space wave travels as a direct ray from the transmitting antenna to the receiving antenna. The other is a ground-reflected ray.

The ionosphere consists of layers of rarified gas that are ionized and made electrically conducting by ultraviolet radiation from the sun. The principle layers are designated E and F. In daytime, the F layer divides into the F_1 and F_2 layers. Under certain conditions these layers reflect radio waves back to the earth. The critical frequency is the highest frequency that will be sent back to the earth when the waves are directed vertically upward. The critical angle is the largest angle with respect to the earth that will cause reflection back to the earth. The maximum usuable frequency is the highest frequency that can be used for sky-wave communication between two points.

Ground-wave communication is reliable for communication between two points, but must use relatively low frequencies and much power. Sky-wave communication is erratic, but great distances can be covered with small amounts of power at high frequencies. Interference from man-made and natural sources hinder radio reception. Fading also occurs, but can be reduced in part by automatic volume-control circuits.

An antenna has an input impedance, the magnitude and angle of which depend on the frequency and the antenna length. The input impedance of a typical vertical insulated quarter-wave antenna driven across the base insulator is about 36 ohms resistance; at the frequency at which it is one-half wavelength long, the input impedance is about 300 ohms resistance. The input impedance of a half-wave antenna driven at the center is 73 ohms when in free space, and approximately this value when near the earth.

The radiation resistance of an antenna is the power radiated divided by the effective input current squared. The radiation efficiency equals the power radiated divided by the power input. The antenna resistance equals the power supplied divided by the square of the input current.

Broadcast antennas for amplitude-modulation systems often are vertical towers that are nondirectional, but sometimes two or more towers are used to provide directional radiation patterns. The patterns are determined by finding the radiation produced at various points by each antenna and then combining the individual radiation to find the resultant strength.

For point-to-point communication, directional antennas are used. These may be composed of half-wave antennas arranged in tier arrays, or may be long-wire antennas, such as the rhombic. Parasitically driven radiators reflect energy that otherwise would be lost out the back of the antenna. For ultrahigh and superhigh frequencies, metallic reflectors are used.

At frequencies above about 30 megacycles, depending on conditions, communication over line-of-sight path is possible. The actual distance is some-

what greater than the optical distance because of refraction, or bending, of the rays by the lower atmosphere.

Antenna feeders are of the mismatched type not terminated in their characteristic impedance and having standing waves on them, or they are of the matched type with no standing waves. A mismatched line is one that is connected directly to the antenna, and is matched to the transmitter with an impedance-transforming circuit so that the transmitter is properly loaded. A matched line is one that is connected to the antenna through an impedance-transforming circuit in such a way that standing waves do not exist to an appreciable extent.

Coaxial cables are used extensively for feeding broadcast antennas. These usually are operated as matched feeders with negligible standing waves.

REVIEW QUESTIONS

1. Explain why the two wires of a transmission line do not radiate appreciable power but the two wires of a half-wave antenna do radiate power readily.
2. Briefly discuss the difference between the radiation field and the induction field.
3. Why does an antenna in free space perform differently than it does when located elsewhere?
4. What is the free-space pattern of an antenna viewed from the end? From the side?
5. What is meant by the term "image antenna"? Why is this viewpoint useful?
6. What is meant by the term "ground-reflection factor"?
7. Why does the height of an antenna above the earth affect the radiation pattern?
8. Discuss the components that comprise the ground wave.
9. What is meant by a sky wave?
10. Which characteristics of the ionosphere are of importance in radio?
11. What is the critical frequency? What happens to waves above this frequency?
12. What is the critical angle? What happens if it is exceeded?
13. What is meant by maximum usable frequency?
14. Discuss the important characteristics of ground-wave transmission.
15. Discuss the important characteristics of sky-wave transmission.
16. What type of transmission is most useful for local broadcast? In long-distance point-to-point service?
17. Why is it possible for a weak received signal strength to be satisfactory in a rural area but not in a city area?
18. What causes fading?
19. Approximately what are the input impedances of the following: A quarter-wave vertical insulated antenna driven across the base insulator? A half-wave vertical insulated antenna driven across the base insulator? A center-driven half-wave antenna in space?
20. Fully explain each of the separate losses that occur in an antenna.
21. What types of antennas are used for ordinary broadcast service?
22. What types of antennas are used for point-to-point service?

23. Briefly describe communication by line-of-sight path.

24. What is the fundamental difference between a mismatched and a matched feeder?

25. Briefly explain the theory of a matching stub.

PROBLEMS

1. Use Eq. (126) and verify the shape of the radiation pattern of the full-wave antenna of Fig. 262.

2. Following the method in the illustrative problem starting on page 499, check the shape of the ground-reflection factor diagram of Fig. 266.

3. Verify the shape of the radiation pattern of Fig. 271 by making calculations for 20°, 40°, and 60°.

4. The antenna whose characteristics are shown in Fig. 284 is to be operated at 590 kilocycles. Write the algebraic expression for its input impedance. If the power input to the antenna is 1.0 kilowatt, what will be the input current? What will be the voltage across the base insulators?

5. Referring to Prob. 4, design a simple impedance-matching network to match the antenna to a 77-ohm coaxial cable. What will the current in the cable be, and what will the voltage between the center and outer cable conductors be? Assume that the units used in the matching network are lossless.

6. Following the method of the illustrative problem starting on page 530, determine the radiation pattern if the magnitudes of the two antenna currents are the same.

CHAPTER XIV

RADIO RECEIVERS

The radio receiver is the final electrical device in a system for radio communication. The **radio receiver** is the device that converts the received radio waves into an audible signal.

At the radio transmitter, the original audio-frequency speech, program, or code signals are raised by modulation to high frequencies for radiation and transmission as electromagnetic radio waves through space. At the distant radio-receiving location, the receiving antenna extracts a small amount of energy from the passing radio wave. The radio-receiving set amplifies the feeble received signal and reduces it in frequency by demodulation (or detection) so that the original signals are audible when a telephone receiver or a loudspeaker is actuated.

Types of Radio Receivers.—As has been explained in the preceding chapters, there are several types of radio communication, depending on the nature of the original signals and on the type of modulation used. Thus there are radio-telephone receivers for speech and program signals, radio-telegraph receivers for code reception, and television receivers. From the standpoint of the system of modulation, the common types are amplitude-modulation receivers, frequency-modulation receivers, and television receivers.

Because of its widespread use, amplitude modulation will receive emphasis in this chapter just as it has in the preceding pages. Two types of amplitude-modulation radio receivers have had extensive use. These are the tuned radio-frequency receiver and the so-called "superheterodyne" receiver. Of these, the superheterodyne receiver is used almost exclusively today.

Radio-receiving Antennas.—In radio transmission, a radio-frequency signal voltage is impressed on the transmitting antenna, and this voltage forces a signal current into the antenna. The voltage between the antenna wires, or between the antenna and ground, causes the radiation of an electric-field component. The current in the antenna radiates a magnetic-field component.

These two components acting together contain energy and constitute an electromagnetic radio wave.

If a transmitting antenna with a driving voltage connected will send an electromagnetic radio wave into space, it is logical to expect that a passing radio wave will produce a voltage in an antenna, and of course this is true. If the antenna is properly connected to a radio-receiving set, signal energy will be abstracted from the passing wave. Whether the electric-field component

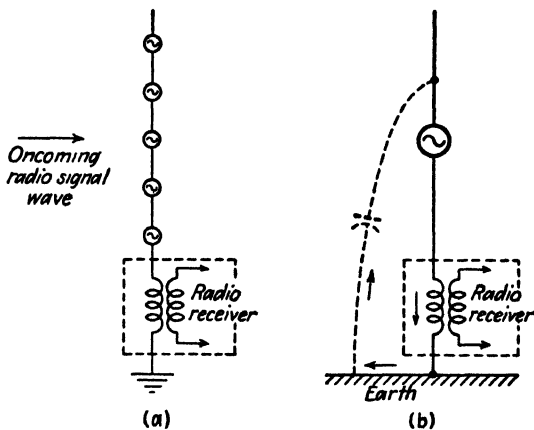



Fig. 304.—It can be considered that the oncoming radio-frequency signal wave induces a voltage in each elemental length of antenna wire. This voltage may be assumed to be concentrated at one point and to force an alternating signal current through the antenna coil of the radio-receiving set as shown by the arrows in (*b*).

induces the signal voltage in an antenna, or whether the magnetic-field component induces the voltage, is sometimes a debatable question.[1] In fact, the matter may be regarded from either viewpoint.

Nondirectional Single "Straight-wire" Receiving Antennas.—If a vertical antenna such as Fig. 304*a* is considered, a passing vertically polarized wave (or any radio wave with a component that is vertically polarized) will induce a voltage in each unit length of the antenna. For convenience, the actual antenna may be considered to be Fig. 304*b*, where the resultant of the individual voltages appears as a voltage at the center. This voltage can be considered to force a current through the circuit, composed of the primary of the antenna-input transformer and the distributed

[1] For an excellent discussion, see H. H. Skilling, "Fundamentals of Electric Waves," Chap. VII, John Wiley & Sons, Inc.

capacitance to ground. This primary current induces a signal voltage in the secondary of the input transformer of a radio-receiving set.

The electrical equivalent of Fig. 304 is shown in Fig. 305, and it is the standard circuit arrangement for testing a radio receiver.[1] The source of testing voltage should be adjustable over the range of frequencies covered by the receiving set. The voltage should be adjustable and be of known magnitude. Also, the source should be such that the output can be modulated over the audible-frequency range. Such a device is called a **standard-signal generator.** The internal impedance should be negligible. The **standard dummy antenna** represents the impedance of an average receiving antenna.

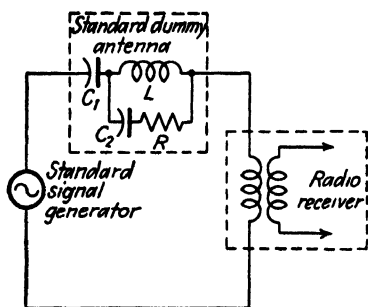



FIG. 305.—Standard dummy antenna used in testing a radio-receiving set. C_1 = 200 micromicrofarads, C_2 = 400 micromicrofarads, L = 20 microhenry, R = 400 ohms.

Directional Receiving Antennas.—The vertical straight-wire antenna of Fig. 304 usually is of random length and essentially is nondirectional, just as the straight vertical transmitting antenna is nondirectional in a horizontal plane near the surface of the earth. In fact, *a given antenna has the same directional characteristics in receiving that it has in transmitting.* This is an application of the **reciprocity theorem** to antennas. In basic terms this theorem states that if a given voltage applied at a point in a circuit causes a given current at another point, the same current will be produced at the first point if the given voltage is inserted at the second point. Because of this reciprocal nature of antennas, antennas in transmission only were considered in the preceding chapter.

Thus any of the directional transmitting antennas, including half-wave antennas, tier arrays, and rhombic antennas, have directional receiving properties, and can be used for directional radio reception. If the directional receiving antennas are remote from the receiving set, as is true usually, transmission lines or coaxial

[1] Standards on Radio Receivers—Methods of Testing Radio Receivers, Institute of Radio Engineers, 1942.

cables must be used. For best results these lines and cables must be matched much as for transmitting.

In special point-to-point systems, where reliable performance is of much importance, very elaborate receiving antenna systems are installed. Sometimes several antennas having slightly different receiving characteristics are used to drive a single radio-receiving set. Thus one antenna may be receiving well when the others are not. Sometimes two or more antennas are located a short distance apart and used to drive the same receiving set, because the signal received at one point may be weak, but it may be strong at

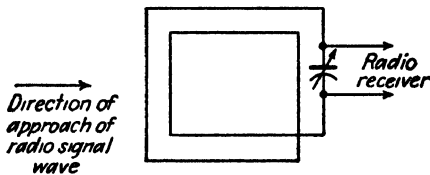



FIG. 306.—Simplified circuit of a loop antenna. Often many more turns of wire are used.

a point a short distance away. Also, slightly different frequencies sometimes are used simultaneously to transmit the same telegraph message, because during certain intervals one frequency transmits well and another does not. These special arrangements are called **diversity systems.**

Loop Antennas.—The loop antenna of Fig. 306 has directional receiving (and transmitting) properties. There are two possible explanations for this, as was mentioned early in the chapter. From the *electric-field viewpoint,* if a radio wave is approaching at right angles *to the plane* of the loop, and if it is considered that the electric component induces the voltages, then *equal* series voltages will be induced in the *same* direction in *each* wire, and will cancel so that no net voltage exists. If, however, the wave is approaching as indicated by the arrow in Fig. 306, equal series voltages will be induced in each wire in the same direction as before; but since the front wires are closer to the distant transmitting antenna than are the back wires, there will be a phase difference between the induced voltages, and, hence, a resultant signal voltage will exist. From the *magnetic-field viewpoint,* if the plane of the loop is parallel to the approaching wavefront, then as the radio wave passed there

could be no change in magnetic flux linkage and no induced voltage. If, however, the wave is approaching as indicated by the arrow, the rapidly changing magnetic-field component will cause the flux linkages to change, and it will induce a radio signal voltage in the loop.

In any event, the loop is tuned to series resonance with the variable condenser for the signal frequency desired, and comparatively speaking, a large current flows around the loop at this frequency. This causes an $E_c = IX_c$ signal-voltage drop across the condenser, and this is the radio signal voltage to be amplified and demodulated in the radio-receiving set. Because loop antennas receive signals from the direction shown in Fig. 306, they are used with radio receivers for direction-finding purposes. It should be mentioned that a loop antenna receives both from the front (as in Fig. 306) or from the back (direction opposite to that of Fig. 306). In fact, its pattern is a figure of eight.

Loop antennas are used extensively in modern broadcast radio receivers because they reduce interference from local electric disturbances, such as electric sparks, etc. The loop is shielded by being placed within a *conducting* metallic tube, or mesh, so that an electric field cannot penetrate and affect the loop. The shield is not entirely continuous, but the path is broken so that currents cannot flow *around the shield.* If this were not done, the demagnetizing effect of the relatively large current that would flow around the shield would render the loop inoperative, in much the same way as a short-circuited transformer turn would render a transformer inoperative. It appears probable that such disturbances as electric sparks from switches, commutators, etc., produce *strong* local (induction) electric fields but *weak* magnetic fields. Hence, the shielded loop, which can be affected only by magnetic fields which will pass through the shield, will receive radio signals by magnetic induction from the wave coming from the distant transmitting station, but the loop will not be acted on by the strong electric fields produced by local disturbances (see also reference listed in footnote on page 551).

The Tuned-radio-frequency Receiver.—Sets of this type formerly were standard for broadcast reception, but now are seldom used. They are relatively simple and have important features; for these reasons they will be considered briefly.

A block diagram of this set is shown in Fig. 307. The radio-

frequency signal voltage induced in an antenna input transformer (Fig. 304) is impressed on a tuned-radio-frequency amplifier (page 551). Most of the selectivity of the set depends on the selectivity of this amplifier. The amplified radio-frequency signal is then impressed on the demodulator, or detector, which distorts the carrier and sidebands and produces an audio component, as explained on page 427. Detectors using triodes were used extensively in early receivers of this type. The grid leak-condenser method (page 435) was an early favorite.

In the tuned-radio-frequency receiver the audio-frequency component produced in the demodulation process is used to drive the

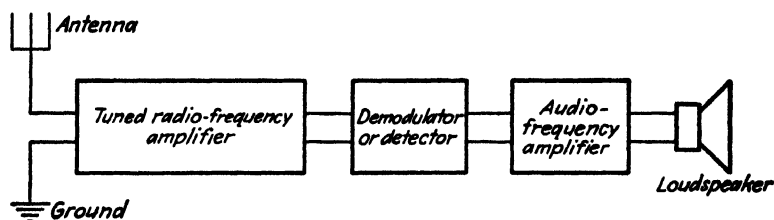



FIG. 307.—Block diagram of a tuned-radio-frequency radio-receiving set.

audio amplifier, which consists of two parts, a voltage amplifier (page 265) and a power amplifier (page 318). The power amplifier drives the loudspeaker.

The Superheterodyne Receiver.—This name throws little light on the electrical principle of operation, but the name is, nevertheless, almost universally used. The name **double-detection receiver** has been employed, and is perhaps more indicative of the method of reception. A block diagram of the so-called "superheterodyne" is shown in Fig. 308, which, as will be explained later, is a little more complete than some sets.

The incoming signal composed of the carrier and the two sidebands is selected and amplified in the tuned-radio-frequency amplifier. This signal is then impressed on a demodulating circuit called by various (and sometimes confusing) names, such as the first detector, the mixer, or the frequency converter. For reasons that will appear later, the purpose of this stage is to reduce the frequency of the signal, as will be explained now.

Suppose that the radio station being received has an assigned carrier frequency of 1,000,000 cycles. If this is being amplitude-modulated by a program occupying the audio-frequency band of

from 50 to 10,000 cycles, the received signal will consist of three components, the carrier of 1,000,000 cycles, the upper sideband of 1,000,050 to 1,010,000 cycles, and the lower sideband of 999,950 to 990,000 cycles. These will be amplified and passed by the tuned-radio-frequency amplifier, and will be impressed on the next unit, which will be called the **frequency converter.** The process is as follows: If the oscillator, or **local oscillator** as it often is called, is oscillating at a frequency of 1,455,000 cycles, and if this frequency and the received radio signal are impressed simultaneously on the frequency converter, sum and difference frequencies will be

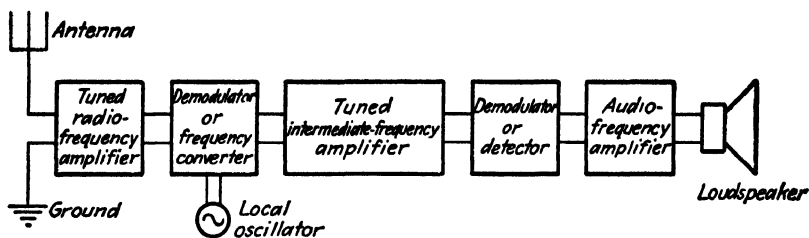



FIG. 308.—Block diagram of a superheterodyne radio-receiving set.

created because this *frequency converter is a demodulating circuit that distorts the waves.* Among the terms created by the demodulator, or frequency converter, are difference frequencies found by subtracting the received radio-signal components from the frequency of the local oscillator. These difference frequencies will be 1,455,000 cycles − (1,000,000 cycles, 1,000,050 to 1,010,000 cycles, 999,950 to 990,000 cycles) = 455,000 cycles, 455,050 to 465,000 cycles, and 454,950 to 445,000 cycles.

Other frequency components also are created in the distortion process in the frequency converter, but these may be disregarded because the tuned intermediate-frequency amplifier will select and amplify *only* the frequencies just listed in the preceding paragraph. These are called the **intermediate frequencies.** They are the complete radio signal containing all the original program variations, but they have been translated or moved by the frequency converter to a lower place in the radio-frequency spectrum. This is called **heterodyne reception,** which is defined as[1] "the process of operation on radio waves to obtain similarly modulated waves of

[1] The definitions given here are from Standards on Radio Receivers—Definitions of Terms, Institute of Radio Engineers, 1942.

different frequency." The reason for reducing the original radio frequencies to intermediate frequencies will appear later.

The **intermediate-frequency amplifier** is a tuned-radio-frequency voltage amplifier of the type described on page 308. Pentodes with tuned transformers (page 312), are often used, and the voltage gain per stage is of the order of 100. The transformers sometimes have air cores with variable parallel trimmer condensers for tuning, or have special "iron" cores that are adjustable and can be used to tune the transformer. In either event, the amplifier selects and amplifies *only* the intermediate-frequencies to which the received radio program is reduced by the frequency converter. The intermediate-frequency amplifier is adjusted (at the factory or by a radio serviceman) to amplify the intermediate frequencies only, and *is not changed by the person operating the radio set.* It is common practice to adjust this amplifier to pass a center frequency of 455,000 cycles, but other center frequencies are used. Note in particular that the adjustments must be such that the sidebands also are passed (see Fig. 173, page 309).

In **superheterodyne reception,** defined as "a form of reception in which one or more frequency changes take place before detection," the signal from the intermediate-frequency amplifier is passed to a demodulator, which usually is a diode detector (page 429). This produces an audio-signal component, and also provides an automatic volume-control voltage. As explained in the discussion of diode detection (page 431), the audio-signal component is amplified in a voltage amplifier, and this increased signal then is impressed on a power tube (usually a pentode) that in turn drives the loudspeaker.

Frequency Conversion.—It is at once apparent that the superheterodyne-radio receiver is somewhat more involved than the tuned-radio-frequency receiver. What is not so apparent is that the superheterodyne gives improved performance, the reasons being evident only after a study is made of the frequency converter and the intermediate-frequency amplifier.

It has been stated that the frequency converter is a demodulator that produces sum and difference frequencies by a process of distortion. It is, accordingly, fundamentally the same as the demodulator described on page 427. The frequency-converter circuit often used will be presented at this time. A discussion of the other

units in the block diagram of Fig. 308 is not necessary, because each unit has been considered in detail elsewhere in this book.

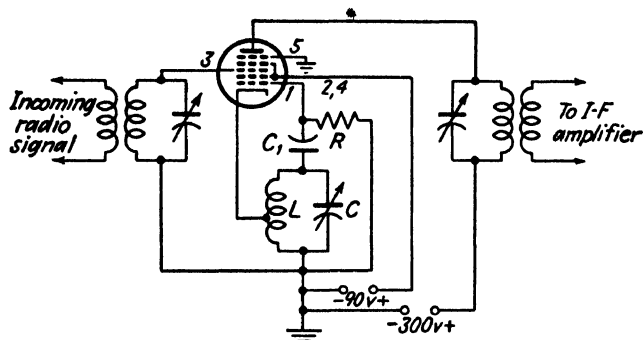
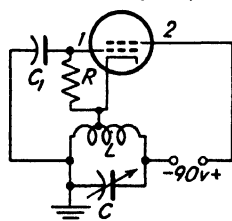
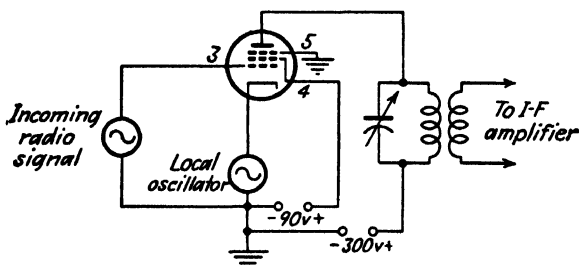
A typical frequency-converter stage employing a type 6SA7 tube is shown in Fig. 309*a*. This tube has *five* grids, two of which are

(a)-Simplified circuit of combined oscillator and frequency converter

(b)-Oscillator portion of (a)

(c)-Demodulator or frequency-converter portion of (a)

Fig. 309.—Demodulator or frequency-converter and oscillator stage of a superheterodyne radio-receiving set.

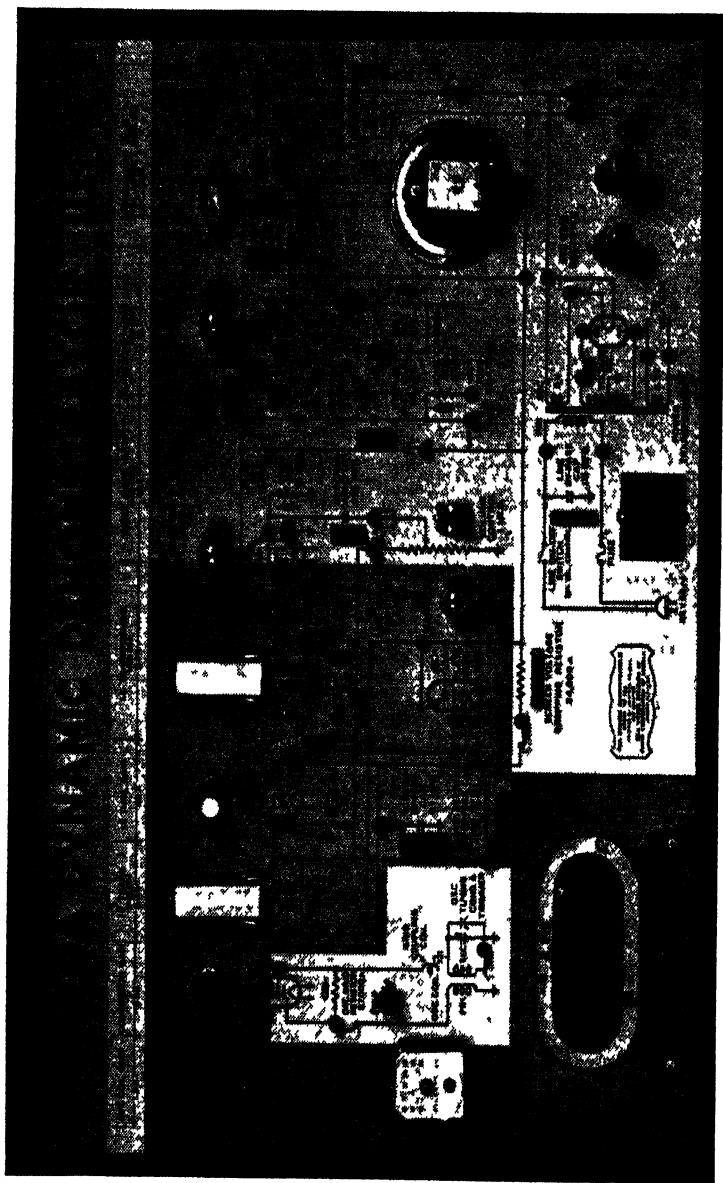
connected together internally. It sometimes is called a **pentagrid converter.** This tube performs two functions: (*a*) It acts as a local oscillator, producing the single frequency necessary to reduce the incoming radio signal to the intermediate-frequency band, and

(*b*) it acts as the demodulator, or frequency converter, to reduce the received radio signal to the intermediate frequencies. This tube should be regarded as *two tubes in one envelope.* For convenience the grids have been numbered 1 to 5 in the usual order. The actual circuit details have been altered slightly to simplify the explanations.

The Oscillator.—This is composed of grids 1 and 2, the tuned L-C circuit, condenser C_1, and resistor R. As indicated, grid 2, which acts as the plate of a conventional oscillator, is at + 90 volts. This circuit is an adaptation of the conventional Hartley oscillator (page 380) and is reproduced in more conventional form in Fig. 309*b*. The oscillator is self-biased by the direct-current component flowing through R (page 377).

The Demodulator.—This utilizes the pentode portion of the tube, as shown in Fig. 309*c*. The bias is such that distortion occurs in the plate circuit (page 419). The incoming modulated radio-frequency signal from the tuned-radio-frequency amplifier and the single-frequency signal from the local oscillator (Fig. 309*b*) are both impressed in series between the control grid and cathode. The intermediate-frequency components, created by the distortion, are amplified and passed by the tuned intermediate-frequency amplifier.

Frequency Selectivity.—There are at least three important characteristics of a radio receiver; these are (*a*) the selectivity, (*b*) the sensitivity, and (*c*) the fidelity. By **selectivity** is meant the ability of a radio receiver to accept signals of a certain frequency and to reject others. The **sensitivity** of a radio receiver is a measure of the ability of a radio receiver to receive weak signals. This depends on the amount of amplification in the stages preceding the final demodulator, or detector, stage, where the audio component is created. The **fidelity** of a radio receiver is a measure of its ability to reproduce faithfully the received message or program. Because sensitivity and fidelity depend to a great extent on amplifier characteristics that have been covered in preceding chapters little additional information will be given. On the other hand, frequency selectivity has not been considered before, and will be discussed at this time.

In the discussion on page 556, it was shown that if a radio-program signal consisting of a 1,000,000-cycle carrier, an upper sideband of from 1,000,050 to 1,010,000 cycles, and a lower sideband

This illustration shows the circuit and parts of an amplitude-modulation superheterodyne radio-broadcast receiver. The equipment is mounted on a panel for instructional and test purposes. The performance of this "Demonstrator" is comparable to an actual radio-broadcast receiver. (*Radio Corporation of America.*)

of from 99,950 to 990,000 cycles were impressed on the demodulating frequency converter with a 1,455,000-cycle signal from the local oscillator, the radio-program signals would be moved down to a carrier of 455,000 cycles, an upper sideband of 455,050 to 465,000 cycles, and a lower sideband of 454,950 to 445,000 cycles. These intermediate frequencies, *and these only*, will be amplified appreciably by the intermediate-frequency amplifier, which has characteristics such as those in Fig. 173, page 309.

In Fig. 308, the main condenser in the tuned-radio-frequency stage and the main condenser in the local oscillator of a superheterodyne are mechanically connected, so that turning one condenser also turns the other. Thus when the local oscillator is oscillating at 1,455,000 cycles so as to react with the incoming signal having a 1,000,000-cycle carrier and produce components that will pass through the intermediate-frequency amplifier, the tuning of the tuned-radio-frequency stage is made to pass the incoming signal having a 1,000,000-cycle carrier.

If it is desired to receive another station having, for instance, a carrier frequency of 1,010,000 cycles, then the dial of the radio receiver is turned to this frequency marking, and this causes the local oscillator to generate a frequency of 1,465,000 cycles. The frequency converter will then produce a difference frequency of 455,000 cycles, and this will pass through the intermediate-frequency amplifier. The discussion in this paragraph neglects the sidebands, but these would be handled along with the carrier, as explained previously.

Although the tuned-radio-frequency stage is tuned each time to the frequency of the station to be received, it cannot completely prevent the signals of unwanted stations from entering the input of the frequency converter. The signal from the local oscillator also will react with each of these unwanted signals, and will produce sums and differences. It is important to note, however, that because of the high selectivity of the intermediate-frequency amplifier, only the wanted signal is received, because this amplifier will pass only that signal which combines with the signal from the oscillator and which produces a signal centered at 455,000 cycles.

To understand one advantage of the superheterodyne, suppose that the set is adjusted to receive a station having a carrier of 1,500,000 cycles. For this purpose the main tuning dial would be set at 1,500,000 cycles, and the local oscillator would produce

a frequency of 1,955,000 cycles to reduce the carrier to 455,000 cycles so that the signal would pass through the intermediate-frequency amplifier. Now suppose that an *unwanted* station having a frequency of 1,520,000 cycles is passing through (to some extent) the tuned-radio-frequency amplifier. This will be reduced by the converter to a frequency of 475,000 cycles. The difference between the two original carriers in the input to the frequency converter is (1,520,000 − 1,500,000)/1,500,000 = 0.013 = 1.3 per cent. But, after frequency reduction in the converter, the difference between the wanted and unwanted signals would be (475,000 − 455,000)/455,000 = 0.044 = 4.4 per cent. This is important, because it means that at intermediate frequencies it will be easier to make circuits that will accept one radio program and reject another close it it. On a percentage basis the converter spreads the wanted and unwanted signals farther apart.

With a superheterodyne, to tune in a wanted station the intermediate-frequency amplifier, which produces much of the gain, *does not need to be varied.* This makes it easier to design. It is difficult to design one tuned circuit that will be equally selective over the entire broadcast range of from about 550 to 1700 kilocycles. But a very selective amplifier can be designed *for a single frequency band,* such as the band for the intermediate frequencies. There is no corresponding problem in designing an adjustable local oscillator that can be varied so that the desired station will be received.

Image Frequencies.—As has been pointed out, a tuned-radio-frequency amplifier is connected between the antenna and the frequency converter. It has been explained that the tuning of this is varied by the same control that changes the frequency of the local oscillator. When the local oscillator is producing the proper frequency to react with the signal from a given station and to produce a component that will pass through the intermediate amplifier, the tuned-radio-frequency amplifier must at this same time be passing the desired signal to the converter. By its selective action the tuned-radio-frequency amplifier assists in obtaining selectivity.

This amplifier is effective also in suppressing the signals of frequency that are known as "image frequencies." An **image frequency** is a signal of frequency which is higher than the frequency of the local oscillator, and which will react in the converter with

the local oscillator to produce a signal component which will pass through the intermediate-frequency amplifier. Thus, if the frequency of the local oscillator is 1,455,000 cycles, a *desired* signal of 1,000,000 cycles will react with this in the converter and produce a 455,000-cycle component that will pass through the intermediate-frequency amplifier. Also, if an *undesired* transmitting station at 1,910,000 cycles produces a signal in the converter, then this will produce a component at 1,910,000 − 1,455,000 = 455,000 cycles that will cause the signal of the undesired station to be reproduced. The 1,910,000 station operates at an *image frequency*.

It also is possible for the signal of an unwanted transmitting station to modulate the signal of a desired station and to produce the effect that is called **crosstalk** or **cross modulation.**

The tuned-radio-frequency amplifier in a superheterodyne suppresses to a degree all signals except the one desired, and, hence, reduces the effect of image frequencies and cross modulation. It also assists in preventing the local oscillator from acting like a miniature radio transmitter and helps in keeping it from radiating a signal out over the receiving antenna.

Throughout this discussion it has been stated or implied that a tuned-radio-frequency stage of amplification is used. Often this is true, but in small radio receivers this usually is a tuned circuit without an amplifier tube.

Miscellaneous Features of Radio Receivers for Amplitude-modulated Signals.—The radio receivers considered in the preceding pages were for *broadcast* reception of amplitude-modulated waves. Certain special receivers, sometimes called **communication receivers,** contain various features not incorporated in those for broadcast reception. Some of these special features will be considered now.

Beat-frequency Oscillator.—A superheterodyne may be equipped to receive continuous-wave telegraphic code signals. These usually are transmitted and received as "spurts" of single-frequency energy (page 474). To make their reception satisfactory it is necessary to create an audio-frequency component for the following reason: A single-frequency wave received by a superheterodyne would be reduced by the converter to a single-frequency wave that would pass through the intermediate-frequency amplifier. This single-frequency wave would be converted by the diode detector into a direct-current component that would start and stop as the

radio-frequency signals started and stopped. These variations would produce audible sounds (clicks) in the loudspeaker, but they would be unsatisfactory for reception. If a second local oscillator known as a **beat-frequency oscillator** injects a single frequency into the circuit at some point, such as into the intermediate-frequency amplifier just prior to the diode detector, then modulation will occur in this detector (modulation and demodulation are the same fundamental process of distortion, see page 427) and a tone will be produced by the loudspeaker. Thus if the intermediate frequency is 455,000 cycles, and a single frequency of 454,000 cycles from the beat-frequency oscillator is fed in, a 1000-cycle tone will result. Since the intermediate-frequency signal will be controlled by the "spurts" of incoming radio-telegraphic code signals, the 1000-cycle tone will start and stop in accordance with the code signals to be received.

Crystal Coupling.—For code reception, only a single frequency need come through the intermediate-frequency amplifier. If all other components are eliminated from this amplifier, then less interference will exist in the loudspeaker or the head receiver. To accomplish this, a quartz crystal often is arranged so that it can be inserted into the intermediate-frequency amplifier, and hence, tune this amplifier so that it will pass one very narrow frequency band, and greatly attenuate all other frequencies. In this connection it will be recalled that a quartz crystal has the properties of a resonant circuit (page 383). Provisions are made for removing the crystal or for short-circuiting it when it is desired that the intermediate-frequency amplifier pass a band of frequencies, as it must do in speech or program reception, where the carrier and sidebands are present.

Noise-reducing Methods.—Communication receivers, and receivers for such purposes as transoceanic service, are provided with special arrangements for reducing noise. Among the successful methods are (*a*) the **Lamb noise-reducing circuit,** which causes a burst of incoming static or other similar interference to bias a tube to cutoff and render the receiver inoperative for a moment; (*b*) the **Compandor,** which compresses speech at the sending end and expands it to normal at the receiving end, and thus maintains a higher signal-to-noise ratio at all times; (*c*) the **Codan,** which renders the receiver inoperative when tuning between stations. Another arrangement reduces the amplification and the received

noise during silent intervals between speech words so that the noise does not cause a hearing impairment.

Automatic Frequency Control.—Some of the larger broadcast receivers, and certain of those for special communication purposes, are provided with automatic frequency control. This circuit arrangement assists in tuning in a station and holds the radio receiving set on a station once it is tuned to it. The arrangement consists of a frequency-sensitive circuit, or discriminator (page 449), and a reactance tube (page 444). The discriminator develops a direct voltage proportional to the deviation of the intermediate

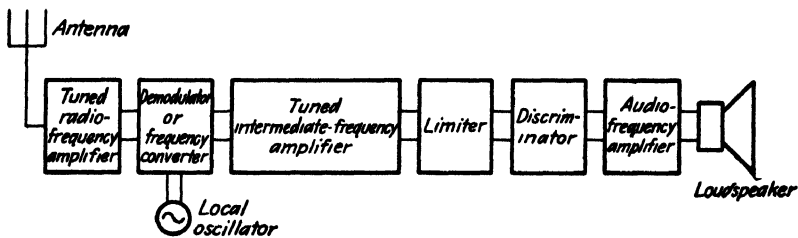



FIG. 310.—Block diagram of one type of radio-receiving set for frequency-modulated signals.

frequency from the desired value, and this direct voltage causes a reactance tube to bring the local oscillator to the correct frequency.

Wave Analyzer.—A wave analyzer for analyzing a steady-state nonsinusoidal audio voltage into its harmonics works on the same basic principle as a superheterodyne radio receiver but operates over the audio range instead of over the radio spectrum. A typical wave analyzer uses quartz-crystal-coupled intermediate-frequency amplifiers to select the various harmonics, and has a vacuum-tube voltmeter to measure the magnitude of the harmonics.

Radio Receivers for Frequency Modulation.—The superheterodyne principle is used in radio receivers for frequency-modulation systems. Special features are incorporated, of course, to make the reception of frequency-modulated signals possible, a block diagram of a typical arrangement being shown in Fig. 310.

The channels used for frequency modulation are in the vicinity of 100 megacycles. During modulation, the frequency deviation is about 75 kilocycles each side of the carrier. The band required by a frequency-modulation system is, however, about 200,000 cycles wide because of the sidebands (Fig. 238, page 442). These statements apply to ordinary frequency-modulation broadcast service.

Because of the very high frequencies used, special antennas are preferable. These must operate satisfactorily over the entire band width. The received signal should be transmitted from the antenna to the receiving set over some type of transmission line or cable. The tuned-radio-frequency amplifier stage performs the same function as in the superheterodyne for amplitude-modulated signals. Likewise, the local oscillator and the frequency converter serve the same purpose. The intermediate frequency to which the frequency-modulated signals are reduced is in the megacycle region instead of at 455 kilocycles as for amplitude-modulation systems. The intermediate-frequency amplifier must pass the 200,000-cycle band.

The **limiter** is a *very important* unit in the frequency-modulation receiver. This device removes all amplitude variations from the frequency-modulated signal. For example, if any electric disturbances, such as lightning or static, cause the amplitude of the received signal to vary, the limiter removes these variations, and permits only the desired frequency-modulated signals to pass through.

The limiter may be a one-stage or a two-stage circuit, a one-stage limiter being shown in Fig. 311*a*. The limiter is a tuned amplifier that saturates and causes nonlinear distortion, as indicated in Fig. 311*b*. The basic circuit is a tuned-radio-frequency amplifier using a pentode. The pentode is operated at a low direct plate voltage of perhaps 60 volts. When the applied frequency-modulated signal has reached a certain value, a further increase in input signal causes no appreciable change in signal output from the limiter. The signal voltage is amplified by the intermediate-frequency amplifier until the voltage is quite high, relatively speaking, and the limiter cuts it down to a low value so that there is little chance for amplitude variations to pass through. Some limiters operate because of grid-current flow, and some utilize limiting action in both the grid and plate circuits. As explained on page 453, if a ratio detector is used instead of a discriminator, a limiter is not needed.

The **discriminator** changes the impressed frequency-modulated signal variations into an output that is the original audio-frequency speech, or program, signal. The action of a typical discriminator was considered in detail on page 450. The audio-frequency amplifier and the loudspeaker are of conventional types, but usually are

of high fidelity because in frequency modulation every effort is made to transmit the audio signals over a wide frequency band.

Sometimes a **preemphasis circuit** is used at the transmitter to accentuate the weak high-frequency components so that they will override noise. If this is done, a **deemphasis circuit** should be inserted in the radio-receiving set between the discriminator and the audio amplifier.[1]

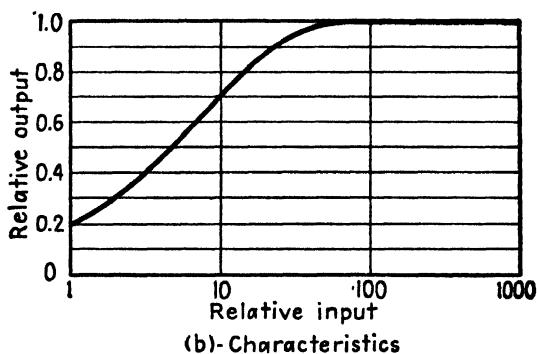
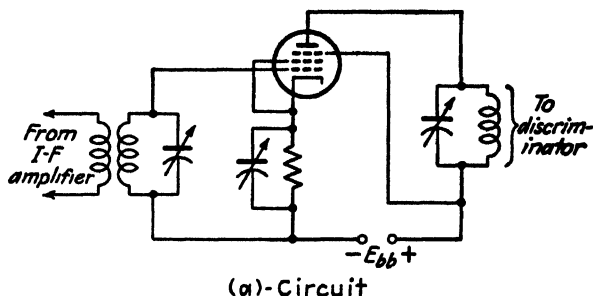



FIG. 311.—Simplified circuit of one type of limiter and its characteristics. This limiter is a radio-frequency amplifier operated so that it saturates, and an increase in the strength of the applied signal does not result in an increase in output. This is indicated in (*b*). (Adapted from reference listed in footnote, page 406.)

Amplitude Modulation versus Frequency Modulation.—Perhaps it is well at this point to compare these two systems. This will be done by briefly discussing the transmitters, the wave transmission paths, the receivers, and noise and interference.

Transmitters.—These were considered in Chap. XII. A brief review indicates that the transmitters used for amplitude modula-

[1] Kiver, M. S., "F-M Simplified," D. Van Nostrand Company, Inc., is recommended to those interested in further information on frequency-modulation receivers.

tion are in general simpler than those for frequency modulation, but perhaps this is because frequency modulation is newer and less standardized. A point not to be overlooked is that the frequency modulation is accomplished at low power levels, using small receiving-type tubes, whereas the opposite is true for at least some methods of amplitude modulation. Attention also is called to the fact that in frequency modulation the power tubes operate at full load at all times, because the modulated signal does not vary in amplitude. This means that extra tube capacity for strong signals need not be provided. Frequency-modulation systems must use channels at very high frequencies because of the wide band needed. The present assigments at the very high frequencies necessitate the use of special tubes and techniques.

Transmission Paths.—As explained in the preceding chapter, the ordinary amplitude-modulation systems operate at the relatively low radio frequencies and utilize both ground-wave and sky-wave paths. Static is particularly bad at such frequencies, and sky-wave paths are erratic. At the very high frequencies assigned to frequency modulation, transmission is essentially by line-of-sight path. Static is less bothersome in this part of the radio spectrum.

Receivers.—The receivers for amplitude-modulated signals are simpler—and hence slightly less expensive, but this probably will not be a permanent condition. The present license requirements are that frequency-modulated signals must be of very high quality. In general, it can be stated that the frequency-modulation receivers are capable of high-quality reproduction. It should be pointed out, however, that either amplitude modulation or frequency modulation *can* provide high-quality signals. Partly because of the received noise, a listener using an amplitude-modulation receiver often turns in the tone control, thus removing the high frequencies from the reproduced program. This is not necessary with frequency modulation because the reception is almost noise-free. But in this connection it will be remembered that evidence exists that the average radio listener does not care for high-quality reproduction (page 20).

Noise.—As previously mentioned, frequency modulation gives almost noise-free reception. It is common knowledge that excellent program reception is possible with an intense thunderstorm in the immediate vicinity. A brief explanation of this is as follows:

Noise can be thought of as being both amplitude-modulated and frequency-modulated. Since an amplitude-modulation system depends on changes in amplitude for its operation, such a system is susceptible to bursts of static, etc., and such interference has never been eliminated in amplitude-modulation reception, although it has been reduced in effect.

The limiter in a frequency-modulation receiver removes all amplitude variations; hence, *amplitude variations* caused by noise and static cannot affect to any great extent the reception of frequency-modulated signals. It does not appear that the frequency deviation of noise signals is great. On the other hand, the frequency deviations of a program signal in frequency-modulation transmitters can be made, and is made, quite large. The magnitude of the instantaneous output of the discriminator depends on the frequency deviation. Hence, if the program has a wide frequency deviation, but noise does not, the signal-to-noise ratio of the output of a discriminator will be high, a very important feature (and possibly the most important one) of frequency modulation. For communication purposes, the frequency deviation used is not so great as for program purposes.

Interference.—In frequency modulation there is less interference between stations, even when they are operating on the *same* frequency. Thus if a good frequency-modulation receiver is located so that it receives signals from two stations on the same frequency, the receiver tends to respond only to the stronger of the two stations. If the stronger of the two signals is about twice as intense as the weaker, then the stronger signal is reproduced and the weaker is suppressed almost entirely.

Television Receivers.—The television pickup tube and the television transmitter were discussed in Chap. XII. It was explained that a television channel requires a very wide frequency band, of at least from 30 to 3,000,000 cycles, to reproduce an image of good quality. Because of this extreme band width, television channels are assigned in the very high frequency region of from about 50 to 100 megacycles, and above.

The electric signals produced when a television camera scans the object, or scene, to be transmitted are used to amplitude-modulate a high-frequency carrier, and are radiated into space. Included with these signals are synchronizing impulses. It is of course necessary that the scanning process and the reproducing process

be in exact synchronism; otherwise the reproduced image will be distorted. The sound accompanying the scene to be televised is transmitted by a frequency-modulation system.

A block diagram of a television receiver is shown in Fig. 312, and it utilizes the superheterodyne principle. A special antenna system must be used because of the high frequencies and the great band width. The radio signals conveying the image, and those transmitting the sound, both are received by the same antenna, and are selected and amplified by the tuned-radio-frequency ampli-

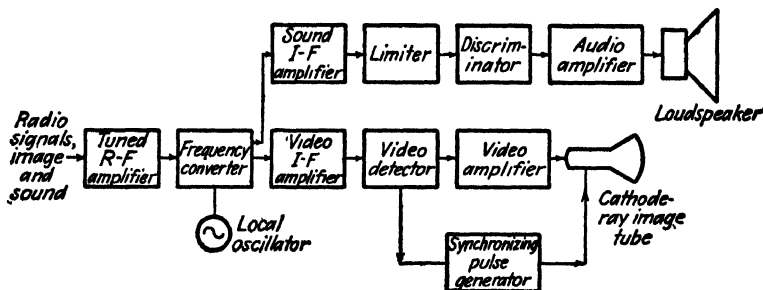



FIG. 312.—Block diagram of a television receiver. The pulses impressed on the cathode-ray tube are to synchronize the electron beam with the scanning in the television camera.

fier and are then converted to intermediate frequencies by the frequency-converter stage.

These intermediate frequencies will be different because the image and sound are radiated at different frequencies at the transmitting station. The intermediate-frequency frequency-modulated sound signals enter the sound intermediate-frequency amplifier, and are then reproduced by the frequency-modulation part of the television receiver.

The amplitude-modulated signals conveying the image enter what may be called the video portion of the television receiver. All these units must be capable of handling the wide band of frequencies present in a television image signal. These signals are amplified, demodulated, amplified again, and impressed on the television image-reproducing tube. The basic principles of operation involved in this process are the same as in the reception of an amplitude-modulated radio program, but the band width is from about 30 to 3,000,000 cycles.

As mentioned, synchronizing impulses are sent out by the television transmitter, and these are utilized at the television receiver to control the action of the image-reproducing tube, keeping it always in synchronism with the television pickup tube.

The image-reproducing tube used in television is a cathode-ray tube, often called a **kinescope.** Before further describing the action of this image-reproducing tube, the basic principle of the cathode-ray tube will be presented.

Cathode-ray Tubes.—There are many types of cathode-ray tubes, but they contain three basic parts: (*a*) an **electron gun** which provides a "pencil," or ray, of high-speed negative electrons which are directed down the axis of the tube; (*b*) either **deflecting plates** or **deflecting coils** which are used to deflect the ray of electrons and thus control the point at which they strike the end of the tube; (*c*) the **fluorescent screen** on the inside of the glass at the end of the tube and on which the impinging electrons produce a visible spot of light. The details of these three basic parts vary widely, and in particular this is true for the tubes used in television. The tubes that are to be described now have been chosen because of their simplicity.

The Electron Gun.—The cross sections of two cathode-ray tubes are shown in Fig. 313. The oxide-coated *cathode* C is indirectly heated. The *control grid* G_1 is of metal and is cup-shaped, surrounds the cathode, and is negative with respect to the cathode. This cup-shaped grid has a hole in it, and because of its negative potential it can control the *number* of electrons that pass down the tube. Control grid G_1 can regulate the intensity of the spot produced on the fluorescent screen. The *accelerator grid* G_2 is a metal disk with a hole at the center, and is made positive so that it will accelerate the electrons. In many cathode-ray tubes accelerator grid G_2 is omitted. The electron beam passes through *focusing anode* A_1, which is a metal tube with two metal disks containing holes at the center. The beam then passes through *focusing anode* A_2, which is a metal cylinder. Anodes A_1 and A_2 are positive with respect to the cathode, but the first anode, A_1, is at a lower positive potential than the second anode, A_2. Because of the shapes in which anodes A_1 and A_2 are made, and because of the difference of potential between them, the shape of the electric field between A_1 and A_2 is such that the electron beam, or ray, is focused

on the fluorescent screen at the end of the tube. Although this method of focusing is very common, magnetic focusing by placing the tube at the center of a coil carrying current is used also.

The Deflecting Mechanism.—The beam, or ray, of negative electrons that travels toward the fluorescent screen can be deflected by a signal in one of two ways: (*a*) by the deflecting plates shown in Fig. 313*a*, and (*b*) by the deflecting coils of Fig. 313*b*.

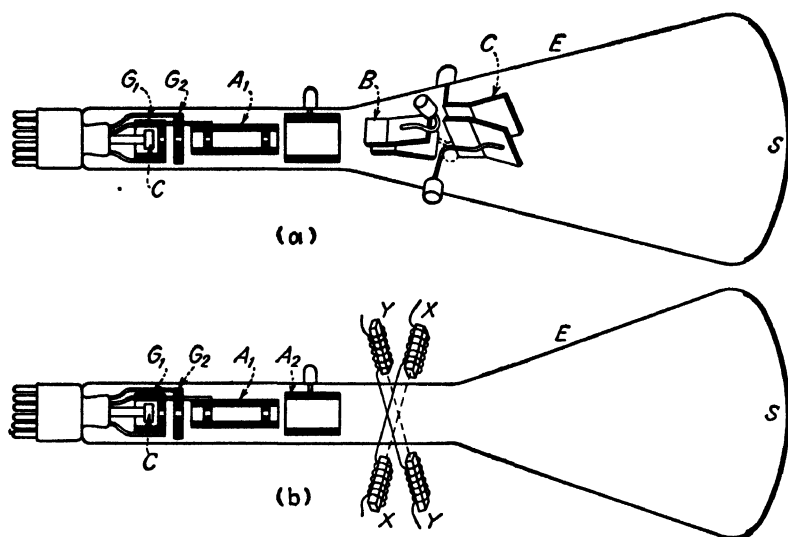



FIG. 313.—Cross section of cathode-ray tubes (*a*) for voltage deflection, and (*b*) for current deflection. In many tubes grid G_2 is omitted. Other electrode shapes are used. The ends of the deflecting plates are shaped as shown so that the deflected electron beam will not strike the ends of these electrodes. (*Radio Corporation of America.*)

The deflecting plates *B* are two parallel metal plates between which the electron beam passes. Deflecting plates *C* are identical with plates *B*, but are at right angles. Because the electrons are negative charges, a voltage impressed between either set of plates will make one plate positive and the other negative, and can be used to deflect the beam. Often two different deflecting voltages are used on the separate sets of plates. The deflecting coils *X* and *Y* of Fig. 313*b* can be used to deflect the beam, because a flow of electrons constitutes an electric current, and an electric current is deflected by a magnetic field. Thus the magnetic fields produced by coils *X* or coils *Y* will deflect the electrons. The two sets of coils can be energized at the same time, and various types of pat-

terns or traces can be produced on the fluorescent screen at the end of the tube. The inner surface of a tube in the region marked *E* usually is coated with graphite. This conducting layer is then connected so that any electrons which strike it may return to the cathode. This prevents islands of electrons from building up on the glass walls of the tube. Such islands of negative electricity would deflect the beam and cause erratic performance.

Television Cathode-ray Tubes.—These are of many types; yet they all operate according to the basic principles just given. Certain important details of a television tube are illustrated in Fig.

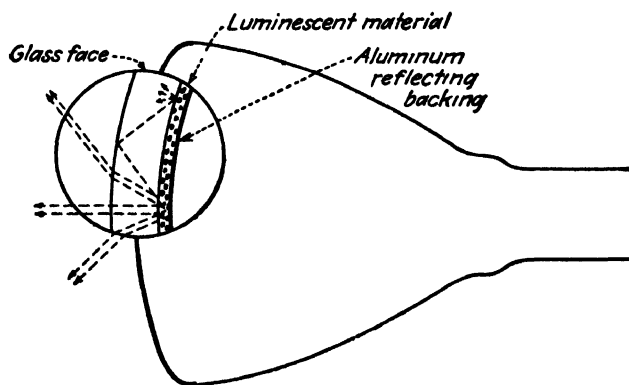



FIG. 314.—A modern cathode-ray television receiving tube with enlarged view showing the light-reflecting action of the thin aluminum backing on the luminescent material. (*American Institute of Electrical Engineers*).

314. The fluorescent or luminescent material on the inside of the end of the glass tube has a *thin* layer, or film, of aluminum as a backing. The electrons coming down the tube *pass through* this thin layer with negligible opposition, and then the electrons penetrate into the fluorescent material (often called a **phosphor**), imparting energy to it, and thus causing it to glow. This thin aluminum backing produces two beneficial effects.[1] (*a*) If it were not for this aluminum backing, some of the light rays emitted by the fluorescent material would travel down the tube toward the electron gun instead of out the end, as desired. This would cause a loss of light energy and a reduction in image contrast, because some of this light would be reflected by the inner part of the tube back again to the fluorescent screen in a random manner. The mirrorlike

[1] Epstein, D. W., and L. Pensak, Improved Cathode-ray Tubes with Metal-backed Luminescent Screens, *RCA Review*, Vol. VII, No. 1, March, 1946.

reflecting aluminum surface immediately reflects the back radiation and redirects it out the end of the tube. (b) In early cathode-ray television tubes the positive ions formed by collision with residual gas particles would get in the beam path, strike the fluorescent material, and cause a yellowish-brown blemish at the center of the screen. These positive ions cannot penetrate the aluminum layer, and hence this discoloration does not occur in the new tube described here.

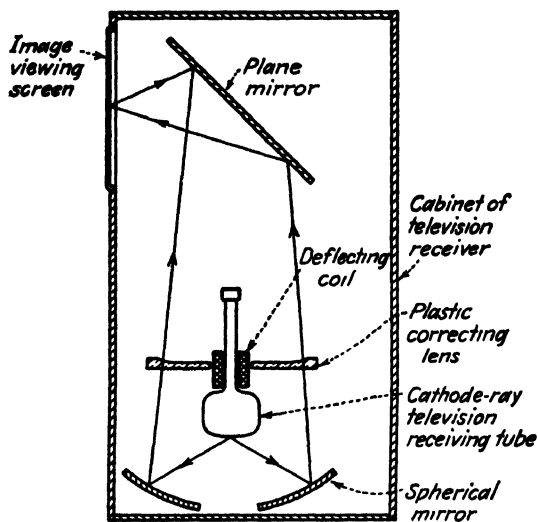



FIG. 315.—Optical system of a modern television-receiving set. *American Institute of Electrical Engineers.)*

Optical Systems in Television Receivers.—The television image reproduced on the fluorescent screen of a cathode-ray tube may be viewed directly. In this event, the image is small unless a tube with a large screen is used. Tubes with screens 20 inches in diameter are available.[1] Another, and for some purposes, a more satisfactory way of obtaining a large image is to use a smaller tube and an optical multiplying system that increases the image size.

One successful method of accomplishing this is the television optical system of Fig. 315. The cathode-ray tube has a fluorescent screen about 5 inches in diameter. The screen produces an image, or picture, too bright for direct viewing. The large spherical mirror beneath the tube directs the light rays upward. A lens of

[1] Kiver, M. S., "Television Simplified," D. Van Nostrand Company, Inc., is recommended to those interested in the details of television.

plastic material is placed between the spherical reflecting mirror and the plane reflecting mirror. This lens corrects optical distortion that otherwise would exist. The plane reflecting mirror reflects the light rays to the relatively large viewing screen.

The Cathode-ray Oscilloscope.—This device, often called a cathode-ray oscillograph, is one of the most useful of all electric instruments. Its uses in radio are so extensive that no attempt will

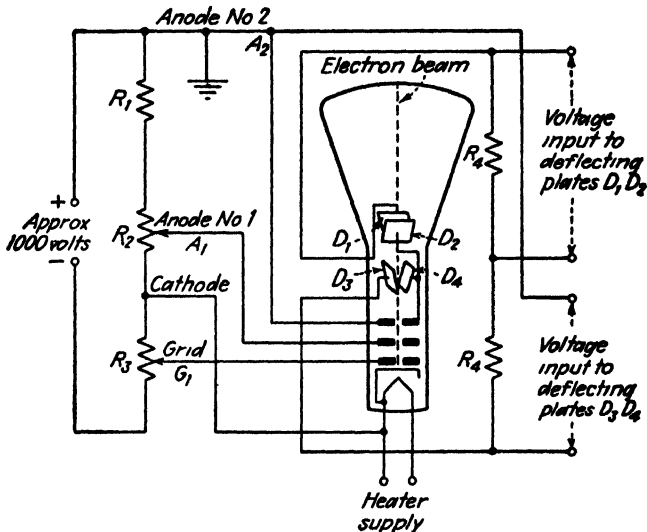



FIG. 316.—Simplified diagram of a widely used cathode-ray oscilloscope. The actual shapes of the electrodes are shown in Fig. 313.

be made to enumerate them. Suffice it to say that any variation in any field of science or engineering that can be converted into a corresponding voltage or current can be studied with a cathode-ray tube. Whether the device is an oscilloscope or an oscillograph may be argued, but it seems reasonable to use the term **oscilloscope** when the reproduced trace on the screen is viewed directly, and to use the term **oscillograph** when the trace is photographically recorded.

The basic circuit of one of the simplest and most satisfactory oscilloscopes is shown in Fig. 316. It will be noted that the accelerating grid G_2 is not used in this tube; otherwise the tube functions as explained on page 571. In much vacuum-tube work the negative cathode is grounded, and the plate, or anode, is at a posi-

tive potential above ground. In the circuit of Fig. 316, the positive terminal is grounded, and the cathode is at a high negative potential with respect to ground. Voltages of from several hundred to several thousand volts are used, around 1000 to 1500 volts being common. Extreme care should, accordingly, be exercised in working with such circuits. The voltage of grid G_1 can be varied, as indicated, to control the *intensity* of the beam. The voltage of

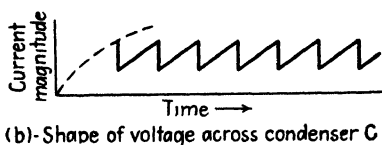
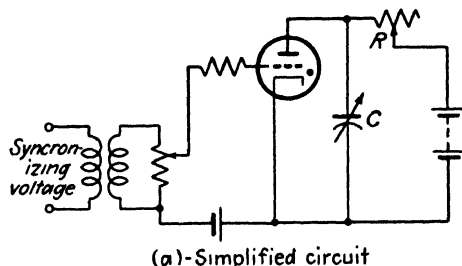



FIG. 317.—The "saw-tooth" or sweep voltage used on a cathode-ray tube is generated by a relaxation oscillator of the form shown, using a gas tube. The condenser C charges slowly as determined by resistor R. At a certain voltage (page 205) the gas tube breaks down, and the condenser discharges through it. This discharge is very rapid. The tube then ceases to conduct, the condenser starts to charge, and the cycle is repeated. To make the image remain stationary on the screen, a small amount of the signal voltage under study is injected through a synchronizing circuit. This voltage is between the grid and cathode and trips or fires the tube just before the condenser voltage rises to the critical value.

anode A_1 can be altered so that the potential between anodes A_1 and A_2 can be adjusted to the correct value to focus the cathode-ray beam and to produce a "sharp" trace on the fluorescent screen.

For viewing signal-voltage variations that are repetitive and occur essentially the same, time after time, a **sweep voltage** is put on one set of deflecting plates to produce a zero axis. This sweep voltage is connected between the two *vertical* plates that produce a *horizontal* motion of the beam. Sweep voltages often are produced by a form of oscillator (Fig. 317) using a gas triode, or Thyratron (page 204). The sweep voltage increases gradually and (relatively speaking) slowly moves the beam across the fluorescent screen, causing a trace. The voltage falls to zero in an extremely

short time and returns the beam to the original position so fast that no trace is left on the screen. With this "saw-tooth" voltage causing a horizontal deflection and the signal voltage to be studied causing a vertical deflection, the signal-voltage variations are "stretched out" and made visible so that they can be studied to ascertain if circuits are operating correctly. This action is illustrated in Fig. 318.

A current may be studied by passing the current through deflecting coils. Since most oscilloscopes are arranged for electric-field deflection using impressed voltages on the deflecting plates, it is a simple matter to pass the current to be studied through a resistor and to impress the corresponding voltage-drop variations on the deflecting plates.

Fig. 318.—If an alternating signal voltage is applied to the vertical deflecting plates of a cathode-ray tube, the electron beam will move up and down producing trace *A-B* on the end of the tube. If the "saw-tooth" wave (*b*) is placed on the horizontal deflecting plates, the electron beam will be swept slowly across the screen as the voltage wave builds up, and the beam quickly will return when the voltage rapidly drops. The beam returns so rapidly that it has insufficient time to produce a return trace. By the action of the saw-tooth sweep, line *A-B* is stretched out and the alternating voltage under observation produces the trace *C-D*.

The frequency of an unknown voltage or current can be determined with an oscilloscope in a simple way: A calibrated oscillator of known adjustable frequency is connected to one set of deflecting plates. A beat-frequency oscillator (page 425) is excellent for this purpose. A voltage corresponding to the voltage or current to be studied is connected to the other set of deflecting plates. When the two signals are of the same magnitude, the same frequency, and *in phase*, the trace on the screen will be a straight line at 45°, as Fig. 319 indicates. If the two voltages are 90° out of phase the trace will be a circle. This method assumes that a continuously adjustable beat-frequency oscillator is available so that the oscillator can be adjusted to the unknown frequency. This is the most simple procedure, and it is recommended for use. However, an oscillator that is not continuously adjustable, or that does not have the necessary range, can be used and what are known

as Lissajou's figures will indicate the frequency ratio. The use of these figures is explained in instruction manuals on the use of the oscilloscope.

The *phase* difference between two signals can be found with the arrangement of Fig. 320. Suppose that it is desired to find the

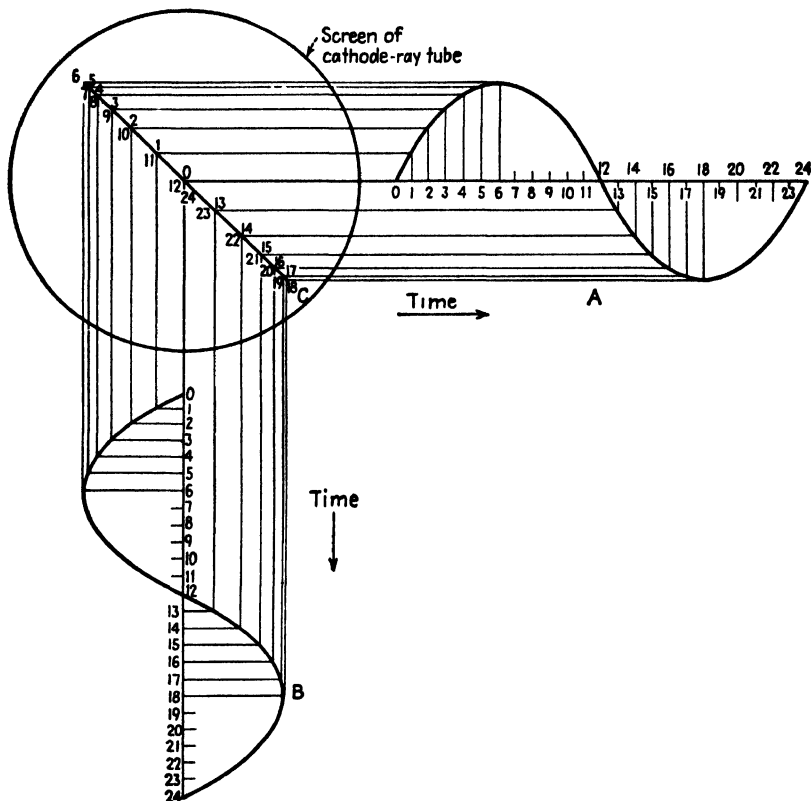



FIG. 319.—The shape of the trace made on the screen of a cathode-ray tube can be determined by projecting points on the impressed signals. For the instant that time $t = 0$, the magnitude of each wave is zero, and hence the beam will be at the center of the screen. At time $t = 6$, both waves are at maximum values, the beam will have two equal forces acting at right angles, and will be deflected to point 6 as shown. Thus when two waves A and B have the same frequency, magnitude, and are in phase (positive on the left), trace C results. This can be proved by analyzing the position at other times such as $t = 12$, 18, and 24.

phase difference between the voltages at two different points in a circuit, for instance, the input and output voltages of an amplifier. From this can be determined the delay distortion (page 267) of the amplifier. If one set of deflecting plates is connected to the input voltage, and the other set is connected to the output voltage,

a straight line will result if the voltages are in phase, a circle will result if they are 90° out of phase, and an ellipse will be produced by any angle in between. To evaluate this phase angle, one set of plates is connected to a phase-shifting network, and then the network is so adjusted that the trace is a straight line. From the values of resistance and capacitance needed to produce the in-phase condition, the phase angle can be computed.

Aligning Radio-receiving Sets.—As has been stressed in discussing the intermediate-frequency amplifier, the characteristics of this amplifier are very important in determining the performance of the radio-receiving set. These amplifiers are factory adjusted and usually maintain these adjustments for long periods. Because of improper handling, or for other reasons, it is sometimes necessary to readjust these amplifiers. A cathode-ray oscilloscope is useful for this purpose.

FIG. 320.—Showing how a cathode-ray oscilloscope can be used to measure the phase shift of a circuit such as an amplifier. From the values of C and R required to produce a straight-line trace, the phase angle can be computed. If necessary, C and R may be interchanged.

In making adjustments in this way, a special signal generator is used. This generator produces a frequency-modulated signal that varies about 30,000 cycles each side of the center frequency. These variations occur at a definite rate. This frequency-modulated signal is impressed on the input of the intermediate-frequency amplifier of the radio set under test, and one set of oscilloscope deflecting plates is connected to the output of the demodulator, or detector (Fig. 308). The other set of deflecting plates is connected to the signal generator to the voltage that is causing the frequency-modulated wave to vary at a definite rate.

With these connections, the cathode-ray beam is caused to sweep across the tube in synchronism with the frequency shift of the generated frequency-modulated test signal. The center frequency of this signal is the same as the center frequency of the amplifier. At an instant when the frequency-modulated signal is, at, say, −30,000 cycles, the beam will be at one side of the oscilloscope tube. When the signal is on the center frequency, the beam will be at the center of the tube. When the signal is at +30,000 cycles, the beam will be at the other side of the tube.

The output of the intermediate-frequency amplifier and of the detector will at each instant be determined by the adjustment of the amplifier and the frequency at that instant. Thus the image that appears on the end of the tube is a plot of the frequency-amplitude characteristics of the amplifier. When this image is visible, it is then possible to adjust the trimmer condensers or magnetic cores of the tuned transformers of the intermediate-frequency amplifier until the desired characteristics are obtained, as indicated by this image curve. For proper adjustments (Fig. 173, page 309) the curve should not be too broad, or undesired frequencies will pass through the amplifier; neither should the curve be too peaked, or the sidebands will not be amplified properly.

SUMMARY

The radio receiver is the final electric device in a radio system. It amplifies the received radio signals and reduces the frequency of the speech, program, or code variations until they exist at the frequencies at which they originally were produced.

The passing radio wave induces a feeble radio-frequency signal voltage in the receiving antenna, and signal energy is abstracted from the radio wave. This causes a minute current to flow in the input transformer and a radio-frequency signal voltage to be induced in the secondary.

Nondirectional single "straight-wire" antennas of random length are suitable for broadcast reception, where reception at many different frequencies is desired. For point-to-point service directional antennas are used. The directional transmitting antennas of the preceding chapter have directional receiving properties in accordance with the reciprocity theorem. The rhombic antenna is used extensively in receiving. The loop antenna is a directional antenna quite common in radio receivers for broadcast reception and direction-finding apparatus.

The type of radio receiver most widely used is known as the "superheterodyne." Tuned-radio-frequency radio receivers are used to a limited extent. All good superheterodynes contain at least a tuned-radio-frequency circuit just before the converter, and many have an amplifier tube in this circuit. This circuit suppresses image frequencies.

In the tuned-radio-frequency amplifier, the received radio signals are amplified in tuned-radio-frequency amplifiers, and then demodulated. For receiving broadcast programs which exist at many frequencies, the radio-frequency amplifiers must be adjustable. This is a limitation because it is difficult to design amplifiers that work equally well over a wide range.

In the superheterodyne the frequency range of the amplifiers is not varied in receiving a program. Instead, the frequency of a local oscillator is varied, and by injecting this and the signal to be received into the frequency converter, the desired radio message or program is reduced in frequency so that it will pass through the intermediate-frequency amplifiers, where most of the ampli-

fication occurs. The frequency converter is essentially a demodulator. The intermediate-frequency amplifier is a fixed tuned-radio-frequency amplifier. Much of the frequency selectivity of the superheterodyne depends on the characteristics of this amplifier.

Radio receivers for special communication purposes are equipped with a second oscillator to render audible continuous-wave telegraph signals. Also, these contain quartz-crystal coupling circuits to narrow the range of amplification and thus exclude noise and interference. Other noise-reducing methods are employed in special radio receivers. Many radio receivers are equipped with automatic frequency control.

Radio receivers for frequency modulation are superheterodyne receivers with a limiter and a discriminator added. The limiter is an amplifier that distorts the intermediate-frequency signals by reducing the magnitude of all signals to a common value. With proper adjustment, amplitude variations cannot pass through the limiter. This reduces noise. The discriminator changes the frequency-varying signals to an amplitude-varying audio signal. If preemphasis is used at the transmitter to raise the weak high-frequency components above the noise level, then deemphasis should be used at the receiver to return the high frequencies to their proper ratio. If a ratio detector instead of a discriminator is used, then the limiter is unnecessary.

The outstanding feature of frequency modulation is that the reception is almost entirely free from noise and interference. Wide-band or high-fidelity programs are broadcast in frequency modulation, but it should be remembered that wide-band programs can be broadcast, and are broadcast, by amplitude-modulation systems. With frequency modulation, the lack of noise makes it more satisfactory to receive the entire frequency band transmitted, instead of using the tone control as is so often done to remove the high-frequency components and the accompanying noise in amplitude-modulation reception.

Television receivers utilize the superheterodyne principle. The amplitude-modulated television image signals are used to operate a cathode-ray tube, sometimes in television called a "kinescope." The signal variations cause the cathode-ray beam to "paint" the television picture image on the end of the tube. Optical systems are used to enlarge the television picture.

A cathode-ray tube consists of three principal parts: (a) the electron gun, (b) the deflecting mechanism, and (c) the fluorescent screen. The cathode-ray tube is used in the oscilloscope, an instrument of wide application.

REVIEW QUESTIONS

1. Enumerate and briefly describe the types of radio receivers.

2. Briefly explain how a radio-receiving antenna functions.

3. What is a dummy antenna, and how and why is it used?

4. What is the reciprocity theorem as applied to radio-receiving antennas?

5. Explain how the shielded loop functions in reducing local interference. Does this apply to static coming from distant sources?

6. What are some of the advantages of the superheterodyne principle of reception?

7. Describe the frequency converter. What basic principle of operation is involved?

8. Explain how a given program is "tuned in" with a superheterodyne.

9. On page 559 it states that the sensitivity of a radio receiver depends on the amount of amplification preceding the final detector. Why?

10. Discuss the relation between coupled-circuit theory (page 107) and the intermediate-frequency amplifier.

11. If a superheterodyne does not have some selectivity preceding the converter, what results?

12. In what respect is a superheterodyne like a small radio transmitter?

13. Why should the automatic volume control be shut off when the beat-frequency oscillator in a communications receiver is used?

14. Briefly describe several noise-reducing methods used in radio receivers.

15. In what respects does a frequency-modulation receiver differ from the usual amplitude-modulation receiver?

16. What are the main features of a frequency-modulation system that make essentially noiseless reception possible?

17. Referring to the discussion on page 565, what is meant by frequency deviation?

18. Enumerate the steps involved in the reception and reproduction of a television image.

19. Repeat Question 18 for the accompanying sound.

20. What is a kinescope, and how is it used?

21. What are the basic parts of a cathode-ray tube, and how does each function?

22. Why is the thin aluminum layer placed on back of the fluorescent screen in certain television tubes?

23. Describe the optical system used in one type of television receiver.

24. Discuss cathode-ray oscilloscopes and their application.

25. Why are saw-tooth waves used in oscilloscopes, and how are these waves produced?

PROBLEMS

1. Calculate the magnitude and angle of the equivalent series impedance of the dummy antenna of Fig. 304 at 0.5, 0.1, and 1.5 megacycles.

2. If a station having a carrier frequency of 550 kilocycles is being received by a tuned-radio-frequency receiver, and a station with a carrier at 570 kilocycles is interfering, what is the percentage difference between the two? If they are being received on a superheterodyne with an intermediate-frequency amplifier of 465 kilocycles, what is the percentage difference between the wanted carrier and the unwanted carrier in this amplifier? At what frequency would the local oscillator be operating?

3. The cathode-ray tube of Fig. 313*b* is arranged so that the *Y*-*Y* coils are vertical and the *X*-*X* coils are horizontal. Direct current is going in the top *Y* coil and out the bottom *Y* coil. Draw a sketch illustrating how the cathode-ray beam will be deflected. Direct current is flowing in the left *X* coil and out the right *X* coil. Draw a sketch illustrating how the beam will be deflected. In the sketch showing the deflections, view the tube at the fluorescent screen end with the base directed away.

4. On page 577 it states that two voltages of equal magnitude, equal frequency, and 90° out of phase will produce a circular trace on a cathode-ray tube. Using the method of Fig. 319, prove this to be true.

5. In the phase-shifting network of Fig. 320 the resistance is 1055 ohms and the capacitance is 0.51 microfarad when the figure on the tube is a straight line. What is the phase shift *in the circuit* under study? Is the angle of the shift leading or lagging? The frequency is 1000 cycles.

INDEX

S